

On-field Respiratory Monitoring Techniques to Track Performance and Exertion of Team-sports Athletes



Runbei Cheng

St Edmund Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Supervised by Prof. Jeroen Bergmann

TT 2026

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Jeroen Bergmann, for his continuous support over the years, in and out of the lab. I would not be here today if it weren't for him coming to my aid time and time again in my hours of need. Thank you for being the best mentor one could've ever asked for.

And of course I would not have even had the opportunity to read for this degree if it weren't for Prof. Martin Bureau who introduced me to Prof. Mark Thompson, who then in turn, introduced me to Jeroen.

A shout out to my amazing lab mates, Dr Vikranth Nagaraja, Dr Emily Kwong, Dr Monica Armengol, Dr Leonardo de Almeida e Bueno, Dr Joseph Russell, Phoebe Haste, and Elyse Levens. Thank you all for your friendship and all the times you have assisted me throughout this degree.

I would also like to thank my housemates Tiff and Owen for taking turns cooking, so that I could spend every two out of three evening leisurely procrastinating.

A toast to all the amazing adventurers that have honoured me with their presence at my DnD table over the years. To Amelia, Nazmus, Odhran, Michaela, Chris, Manos, Eilidh, Lorenz, Tiff, Dan, Niko, and Kori, cheers! And a special thanks to Niko and Kori for sharing my love for food, game nights have been all the more special thanks to you.

And to the Goebels, thank you for welcoming me into your lovely home

time and time again, and for generously loaning me the money I needed when my initial funding plans fell through.

And finally, I would like to dedicate a few words to four very special people in my life. To Manu, thank you for sharing your ambitions with me and always pushing me to be a better version of myself. To Eilidh, thank you for all the times you've turned my tears into laughter, my world is brighter with you in it. To Tiff, thank you for always inspiring me to create new fantastic worlds full of strange stories. And to Lorenz, thank you for being the one I could always talk to about anything and everything. For someone who's estranged from their biological family, I have really luck out with not just one but four amazing chosen families. I would like to thank each and everyone of you for providing with a sense of belonging.

Abstract

Participation in sports has become an essential part of healthy living in today's world. And with advancements in sensor technology and data analytics over the years, many team sports have turned to technology-aided, data-driven, on-field monitoring techniques to help players optimize their games and training. With the adoption of sensing technology, there is a growing interest in capturing fatigue and injury risk for the purpose of preventing injuries and improving overall athlete well-being. Much of on-field sports monitoring aimed at improving athlete well-being can be categorized as either impact monitoring, which aims to capture acute injuries due to physical collisions; or workload monitoring, which aims to capture fatigue and predict injury risk, and can be further subdivided into two categories: external workloads and internal workloads. External workloads are defined as measures of physical work done by the athlete, such as peak speed or distance covered. Internal workloads, on the other hand, are measures of physiological and psychological responses to the external workloads. Through a literature review, we found that the field of internal workload monitoring is dominated by heart rate (HR) related metrics and ratings of perceived exertion (RPE). Meanwhile, respiratory measures, which contain informative biomarkers commonly used in medicine, are underutilized in sports outside of lab settings despite their established physiological relevance. Thus, this thesis presents a body of work on the development and evaluation of a microphone-instrumented mouthguard

designed to capture breathing under sports conditions, and establishes evidence that respiratory data can support perceived exertion estimation. A set of literature reviews was first conducted to identify gaps in current athlete-monitoring practice. A systematic scoping review of on-field monitoring techniques in team sports showed that respiratory monitoring is underused in sports. An umbrella review of wearable respiratory monitoring identified four main sensing approaches: indirect estimation from other physiological signals, chest-wall movement sensing, airflow-property sensing, and acoustic sensing. A rapid evidence review of microphone-based respiratory monitoring in exercise contexts then proved that acoustic sensing is a viable approach to monitoring breathing under sports conditions. Together, these reviews established the need for a robust, unobtrusive respiratory-monitoring system suitable for team contact sports.

A pilot study was carried out to investigate whether respiratory audio could be captured from two microphone locations originally proposed for integration into existing sports protective gear, integrated into a headgear or embedded inside a mouthguard. The oral-cavity microphone produced substantially higher signal quality and enabled more accurate breathing-rate estimation than the headgear placement. Further analysis also found a strong correlation between breathing audio volume and minute ventilation. These findings motivated the development of a custom microphone-instrumented mouthguard, with user comfort and safety as key considerations throughout the development process.

A larger data-collection study was then conducted during Yo-Yo Intermittent Recovery Level 1 tests. Respiratory audio was recorded continuously from the newly developed instrumented mouthguard, while metabolic

data, heart rate, and GPS-derived running speed were collected using a COSMED K5 portable metabolic cart and a Polar H10 heart rate monitor, which are state-of-the-art reference systems. RPE was also collected using cues written on the ground at the test's starting line. This produced a synchronized, multidimensional dataset to aid in investigating the relationships among breathing acoustics, physiological workload, and perceived exertion.

Using this dataset, a breathing-event detection algorithm was developed to estimate breathing rate from breathing audio captured in the oral cavity. Respiratory audio was transformed into mel spectrograms, converted into a respiratory sound envelope, and analyzed using peak/trough detection. Against manual annotations, the algorithm achieved high event-detection accuracy, with a precision of 98.8%, a recall of 98.1%, and a mean absolute timing error of 0.03 s. Compared with COSMED K5 breathing-rate measurements, the audio-derived breathing rate achieved a mean absolute error of 3.0 breaths per minute and a mean absolute percentage error of 7.2%. These findings are comparable to similar studies in the literature that use acoustic respiratory sensing in other wearable form factors.

Following the successful validation of the mouthguard system to collect breathing rate during exercise, the next study examined whether RPE could be estimated from objective cardiorespiratory metrics, and which cardiorespiratory metric has the most impact on the model accuracy. Linear regression, random forest regression, and multilayer perceptron models were trained using respiratory frequency, minute ventilation, VO_2 , heart rate, and participant characteristics. Random forest regression performed best, achieving a mean absolute error of 1.46 ± 1.13 in leave-one-subject-out cross-validation. Permutation feature importance analysis showed

that minute ventilation was the most important predictor of RPE, followed by heart rate, demonstrating the value of respiratory data for exertion monitoring.

Since minute ventilation was identified as the most important cardiorespiratory feature for RPE modelling, the final study of this thesis aimed to measure minute ventilation from breathing audio captured by the instrumented mouthguard. Previously, breathing audio volume was shown to correlate with minute ventilation in the pilot. In this study, deep-learning models were developed to estimate minute ventilation directly from breathing audio. Multilayer perceptron, bidirectional LSTM, and hybrid CNN-BiLSTM models were compared using COSMED K5 minute ventilation as the reference. The hybrid model performed best, achieving an RMSE of 13.50 L/min, MAPE of 9.96%, and r^2 of 0.83 in cross-validation, showing that the instrumented mouthguard can be used to monitor both breathing rate and minute ventilation under sports conditions.

Overall, this thesis demonstrates that a mouthguard-mounted microphone can capture good-quality respiratory information during high-intensity intermittent exercise, that respiratory measures are strongly relevant to RPE estimation, and that deep learning techniques can be applied to extract minute ventilation from breathing audio. These findings extend existing microphone-based respiratory monitoring research from controlled settings toward sport-specific, on-field applications, and position acoustic respiratory sensing as a promising route for future internal workload monitoring in team contact sports.

List Of Publications From This Work

Published Work

Chapter 2 - Cheng, R. and HM Bergmann, J., 2022. Impact and workload are dominating on-field data monitoring techniques to track health and well-being of team-sports athletes. *Physiological Measurement*, 43(3), p.03TR01.

Chapter 3 - de Almeida e Bueno, L., Kwong, M.T., Milnthorpe, W.R., Cheng, R. and Bergmann, J.H., 2021. Applying ubiquitous sensing to estimate perceived exertion based on cardiorespiratory features. *Sports Engineering*, 24(1), p.9.

Chapter 3 - Cheng, R. and Bergmann, J., 2023. A case for an oral cavity-based respiratory rate sensor system. *IEEE Sensors Letters*, 7(3), pp.1-4.

Chapter 6 - Cheng, R., Haste, P., Levens, E. and Bergmann, J., 2024. Feature importance for estimating rating of perceived exertion from cardiorespiratory signals using machine learning. *Frontiers in Sports and Active Living*, 6, p.1448243.

Pending Publications

Chapter 4 - Cheng, R., Haste, P., Levens, E. and Bergmann, J., Internal and External Workload Measurements During a Standardized Intermittent Exercise Fitness Test

Chapter 5&7 - Cheng, R. and Bergmann, J., Measuring Breathing Rate and Minute Ventilation During High-Intensity Exercise Using A Microphone-Instrumented Smart Mouthguard

List of Figures

Figure	Caption
Figure 1.1	An illustration of a multilayer perceptron.
Figure 1.2	An illustration of a convolutional neural network.
Figure 1.3	An illustration of a standard recurrent neural network.
Figure 2.1	Types of monitoring on-field sports monitoring devices based on placement.
Figure 2.2	Wayne state tolerance curve (WSTC). Reproduced from Namjoshi et al. (2013)
Figure 2.3	Article selection flow diagram for the rapid review on respiratory monitoring in sports and exercise using microphones.
Figure 3.1	A STEVAL-STLKT01V1 development board.
Figure 3.2	An example of a data capture session during the pilot study.
Figure 3.3	An illustration of the FFT-based breathing rate algorithm used in the pilot study applied to an audio clip.
Figure 3.4	A simple block diagram of the processing pipeline of the FFT-based breathing rate algorithm used in the pilot study.
Figure 3.5	Examples of waveforms captured by microphones placed in different locations to visual demonstrate the difference in data quality.
Figure 3.6	Bland-Altman agreement plots between predicted and logged breathing rate across microphone placements in the pilot.

Figure 3.7	Correlation plot between VE and the product breathing rate and breathing audio volume.
Figure 3.8	3D-printed TPU enclosure for embedded mouthguard electronics.
Figure 3.9	Concept design for board and battery placement in the mouthguard.
Figure 3.10	Finalized electronics layout for the instrumented mouthguard.
Figure 3.11	Thermoformed EVA mouthguard with embedded electronics and moisture indicator.
Figure 3.12	Summary of self-impression process
Figure 3.13	An example of a stone dental model used to manufacture mouthguards
Figure 3.14	An example of mouthguard fabrication via EVA thermoforming.
Figure 3.15	An example of the finalized version of the instrumented mouthguard.
Figure 4.1	Yo-Yo IR1 test layout.
Figure 4.2	Chalk-marked Borg CR10 RPE scale at the Yo-Yo test start line.
Figure 4.3	Participant wearing monitoring equipment and indicating RPE during a Yo-Yo IR1 test.
Figure 4.4	Example of the GPS speed data captured by the COSMED used to identify the start of the Yo-Yo test.
Figure 4.5	Microphone data completeness test using a computer-generated signal.
Figure 4.6	Detection of COSMED K5 data loss using timestamp drift.
Figure 5.1	Mel-spectrogram & peak detection based breathing-rate algorithm.
Figure 5.2	Distributions of COSMED and audio-derived breathing-rate measurements.

Figure 5.3	Bland-Altman agreement comparison of audio-derived and COSMED breathing rate.
Figure 6.1	Permutation feature importance boxplots for RPE-estimation models.
Figure 7.1	Architecture breakdown of the CNN-BiLSTM hybrid DNN model for VE estimation.
Figure 7.2	Scatter plots and Bland-Altman agreement comparisons of model-estimated and COSMED-measured VE.
Figure 7.3	An example of applying the hybrid model to audio data collected during a full Yo-Yo IR1 test run.

List of Tables

Table	Caption
Table 2.1	A list of methods used to measure internal and external loading in the literature.
Table 2.2	Included review articles on wearable respiratory monitoring in the umbrella review.
Table 2.3	Coding of included review articles by wearable respiratory sensing technique category.
Table 2.4	List of articles included in the microphone-based respiratory-monitoring review.
Table 2.5	A breakdown of study design and applied technology of microphone-based respiratory-monitoring studies.
Table 3.1	Breathing-rate algorithm performance across microphone placements and conditions in the pilot.
Table 4.1	Modified Borg CR10 RPE scale used during Yo-Yo IR1 data collection.
Table 4.2	Example COSMED K5 breath-by-breath output and data-loss indicators.
Table 6.1	Leave-one-subject-out cross-validation results for RPE-estimation models.

Table 7.1 10-fold cross-validation errors and r^2 values for VE estimation models.

Table 7.2 Leave-one-subject-out cross-validation results for VE-estimation models.

Acronyms

BxB Breath-by-Breath.

bpm Beats/Breaths Per Minute.

BR Breathing Rate.

RR Respiratory Rate.

Rf Respiratory Frequency.

VE Minute Ventilation.

VT Tidal Volume.

VO2 Volume of Oxygen Uptake.

RPE Rating of Perceived Exertion.

HR Heart Rate.

Yo-Yo IR1 Yo-Yo Intermittent Recovery
Test Level 1.

GPS Global Positioning System.

ML Machine Learning.

RFR Random Forest Regression.

DL Deep Learning.

ANN Artificial Neural Network.

DNN Deep Neural Network.

MLP Multi-layer perceptrons.

CNN Convolutional Neural Network.

RNN Recurrent Neural Network.

LSTM Long Short-Term Memory.

Bi-LSTM Bidirectional Long Short-
Term Memory.

RMSE Root Mean Square Error.

MAE Mean Absolute Error.

MAPE Mean Absolute Percentage Er-
ror.

STFT Short-Time Fourier Transform.

FFT Fast Fourier Transformation.

Contents

List of Figures

List of Tables

1	Introduction	1
1.1	Benefits of Team Sport Participation	2
1.2	The Health Risks of Poorly Managed Sport Participation	3
1.3	Protective Equipment in Sports	4
1.4	The State Of On-field Monitoring For Better Performance And Safety	6
1.5	Respiratory Signals as Markers of Internal Workload	8
1.6	Overview of Data Science Techniques	9
1.6.1	Audio Signal Processing	10
1.6.2	Machine Learning	11
1.7	Chapter Outlines	17
2	Literature Review	21
2.1	On-field Data Monitoring Techniques to Track Health and Well-being of Team-sports Athletes: A Systematic Scoping Review	23

2.1.1	Introduction	23
2.1.2	Methods	23
2.1.3	Findings and discussions	24
2.1.4	Workload monitoring	26
2.1.4.1	External workload monitoring in contact sports . . .	29
2.1.4.2	Internal workload monitoring in contact sports . . .	32
2.1.4.3	Workload monitoring in throwing sports	35
2.1.4.4	Workload measures and injury rate predictions . . .	36
2.1.5	Impact monitoring	39
2.1.5.1	Miscellaneous collisions monitoring	39
2.1.5.2	Head impact monitoring	41
2.1.6	Further considerations	52
2.1.7	Conclusion	53
2.2	An Umbrella Review of Wearable Sensors For Respiratory Monitoring	54
2.2.1	Introduction	54
2.2.2	Methods	55
2.2.3	Findings & Discussions	55
2.2.3.1	Modulation of other physiological signals	60
2.2.3.2	Chest wall volume changes and movement	61
2.2.3.3	Properties of the air flow	62
2.2.3.4	Acoustic sensing	63
2.2.4	Conclusion	63

2.3	Breathing Monitoring Using Microphones in Sport and Exercise: A Rapid Review	66
2.3.1	Introduction	66
2.3.2	Methods	66
2.3.3	Findings	68
2.3.4	Discussions and conclusion	70
2.4	Chapter Conclusion	73
3	Development And Preliminary Testing of An Oral Cavity Based Respiratory Monitoring Sensor System	75
3.1	Introduction	77
3.2	Sensor Placement & Breathing Rate Detection	77
3.2.1	Methods	78
3.2.1.1	Data Acquisition	78
3.2.1.2	Signal to Noise Ratio	80
3.2.1.3	Breathing Rate Detection	81
3.2.1.4	Analysis	82
3.2.2	Results	83
3.2.3	Discussion	85
3.3	Using Audio Volume for Minute Ventilation and RPE Predictions . .	87
3.3.1	Methods	88
3.3.1.1	Data Acquisition	88
3.3.1.2	Data Processing	88

3.3.2	Results	88
3.4	Instrumented Mouthguard Development	89
3.4.1	Manufacturing Method Selection	90
3.4.2	Battery Selection	92
3.4.3	Ergonomic Design	92
3.4.3.1	Water Ingress Test Of The Updated Design	94
3.5	Mouthguard Manufacturing Process	96
3.6	Conclusion	99
4	Breathing Audio & Internal and External Workload Data Acquired	
	During YoYo-IR1 Tests	100
4.1	Introduction	103
4.2	Methods	105
4.2.1	Recruitment	105
4.2.2	Experimental Setup	106
4.2.2.1	Equipment	106
4.2.2.2	Ratings of Perceived Exertion	107
4.2.2.3	Test Protocol	108
4.2.3	Preprocessing	110
4.3	Technical Validation	114
4.3.1	Microphone Board Data Completeness Test	114
4.3.1.1	Method	115
4.3.1.2	Results	115

4.3.2	COSMED K5 Data Completeness Test	116
4.3.2.1	Data Acquisition	117
4.3.2.2	Analysis	117
4.3.2.3	Results	118
4.4	Usage Notes	119
4.5	Data Records	120
4.5.1	Recorded Modalities	121
4.6	Conclusion	121
5	Breathing Event Detection From Audio Signal	123
5.1	Introduction	125
5.2	Methods	127
5.2.1	Data	127
5.2.2	Audio Processing	128
5.2.3	Data Analysis	128
5.3	Results	130
5.4	Discussion & Conclusion	131
6	Estimating Rating of Perceived Exertion From Cardiorespiratory Signals Using Machine Learning	134
6.1	Introduction	136
6.2	Materials and Methods	138
6.2.1	Participants	138
6.2.2	Data Collection	138

6.2.3	Data Preprocessing	139
6.2.4	Machine Learning Model Selection	139
6.2.5	Feature Importance	140
6.2.6	Statistical Analysis	140
6.3	Results	141
6.3.1	Model Validation	141
6.3.2	Feature Importance	143
6.4	Discussion	143
6.4.1	Limitations	146
6.4.2	Conclusion	148
7	Measuring Minute Ventilation Using an Oral Cavity Based Micro- phone with Deep Learning	149
7.1	Introduction	151
7.2	Methods	153
7.2.1	Data	153
7.2.2	Deep Learning	155
7.2.3	Model Evaluation	157
7.3	Results	157
7.4	Discussion	161
7.4.1	General Discussion	161
7.4.2	Error Analysis	162
7.4.3	Limitations and Future Studies	163

7.5	Conclusion	164
8	Discussions & Conclusion	166
8.1	Thesis Summary & Retrospective Thoughts	166
8.2	Future Work	174
8.2.1	Integrations With Additional Sensors	174
8.2.2	Applications in Sports	175
8.2.3	Future Applications In Respiratory Medicine	176
8.3	Conclusion	178
A	Common Head Impact Severity Measures	179
B	Breathing Rate Logging Programs	182
C	Large Dataset Metadata Tables	184
D	Breathing Rate and Minute Ventilation Plots Per YoYo Test	189
E	VE Algorithms - Leave One Subject Out Results	220
F	Publication Portfolio	228
	Bibliography	231

Chapter 1

Introduction

Team sport participation provides substantial physical, psychological, and social benefits, but it also exposes athletes to risks, including acute injury, overtraining, overuse injuries, and long-term health consequences. While traditional protective equipment can reduce certain forms of impact-related harm, it does not actively monitor the physiological strain experienced by athletes during training or competition. Contemporary player-monitoring systems have begun to address this gap by tracking external and internal workloads, yet internal workload monitoring remains dominated by heart-rate measures and subjective ratings of perceived exertion. Respiratory measures, including breathing rate and minute ventilation, offer physiologically meaningful indicators of exertion but remain underused in applied sports settings.

Here, we discuss the rationale for respiratory monitoring in team contact sports and outline the potential of acoustic sensing, leveraging signal processing and machine learning techniques, to enable unobtrusive on-field monitoring. Following this, we present the development and evaluation of an oral-cavity-based acoustic respiratory monitoring approach capable of supporting data-driven athlete safeguarding in contact sports environments in this thesis.

1.1 Benefits of Team Sport Participation

Physical activity through sport participation plays an essential role in ensuring healthy living across the world. Sport participation can help improve health, prevent crime, and promote social development (Cameron & MacDougall 2000, Schmitz et al. 2004). It also plays a key role in reducing several key challenges we may face in the future, such as obesity and chronic illness (Hills et al. 2011). In particular, participation in team sports has been reported to be especially beneficial when compared to individual activities, due to the social nature of team sports (Pharr et al. 2019, Eime et al. 2013). These benefits have been observed across various age groups and demographics, highlighting the universal advantages of engaging in team-based sports.

Those who engage in team sports often do so for a longer duration and at a higher intensity compared to individual sports. These higher levels of physical activity lead to comparably better outcomes in maintaining a healthy weight and reducing the risk of chronic diseases such as obesity, diabetes, and cardiovascular issues. In addition to the physical health benefits, engaging in team sports strongly promotes good mental health. Participation in team sports is consistently linked to lower levels of depression, anxiety, and suicidal ideation. Studies indicate that team sports can protect against feelings of hopelessness and contribute to higher life satisfaction (Eime et al. 2013). The social nature of team sports fosters a sense of belonging and community, enhances self-esteem, and provides team-sport participants with a structure of social support, all of which are critical factors in promoting mental health and wellbeing (LaRocca et al. 2023, Tahira 2022).

All these benefits underscore the importance of promoting team sports as a vital component of public health initiatives throughout society. However, there are also risks involved when participating in team sports. These risks can be compounded in sports that involve physical contact between players.

1.2 The Health Risks of Poorly Managed Sport Participation

Taking part in team sports can lead to acute injuries and even permanent disabilities, with young adults at their physical prime exhibiting the highest injury rate (Kujala et al. 1995). Sports that involve more frequent and powerful body contact have reported higher overall injury rates, and there is also a clear indication that injury rates are higher during competition than during practice.

On top of accidental injuries, participation in team sports can also lead to health drawbacks when it is poorly managed. Overly intense training sessions and competitions can lead to burnout and overuse injuries. Burnout due to overtraining can result in a decline in physical performance as well as increased susceptibility to physical injuries (DiFiori et al. 2014, Cardoos 2015). It can also potentially lead to both physiological and psychological symptoms, and even affect the immune system (Keaney et al. 2018). Recovery time following overtraining may vary from days to months, depending on the severity. It's been estimated that around 35% of young athletes have been overtrained at least once (Raglin et al. 2000). Overuse injuries have been estimated to account for 54% of all sports injuries reported at university level (Luke et al. 2011), and can lead to a significant loss of time from sports. Severe overuse injuries can even threaten future sports participation (DiFiori et al. 2014).

In the long term, accumulated head injuries can lead to chronic traumatic encephalopathy (CTE) and significantly increase the risk of developing Alzheimer's disease (AD), Parkinson's disease (PD), and possibly amyotrophic lateral sclerosis (ALS) later in life (Ling et al. 2015, VanItallie 2019). This issue is particularly prevalent in high-contact team sports such as hockey, rugby, soccer, and American football. Research indicates that athletes in these sports can sustain hundreds to over a thousand head impacts in a single season, with many of these impacts occurring without

immediate symptoms of concussion (Bailes et al. 2013). With the popularity of these sports, even small percentages of head injuries can translate to millions of affected individuals, highlighting the widespread nature of the problem (Gutierrez et al. 2014).

In addition to these negative health impacts on the athletes themselves, sports injuries also place a huge burden on the healthcare system, affecting not only the individuals involved but also the broader economic landscape. Numerous studies have emphasized that better injury prevention in sports can save local healthcare systems millions annually (Ryan et al. 2019, Marshall et al. 2016, Finch et al. 2015, Ackerman et al. 2019). Moreover, the indirect costs associated with sports injuries, such as loss of income and productivity, are significant. Injuries among working-age individuals lead to substantial healthcare costs and necessitate disability support, which places additional strain on healthcare resources (Axelrod et al. 2022). The economic impact is not limited to direct medical expenses; it also includes the societal costs of lost productivity and the need for rehabilitation services (Vriend et al. 2017).

1.3 Protective Equipment in Sports

These health and financial drawbacks indicate a clear need to develop suitable injury prevention methods and risk management approaches. Existing protective equipment for sports aims to shield players from acute injuries (Figure 1.1 shows an example of hockey players wearing full-body protective gear during a match). For example, in the case of head injuries, there is clear evidence that helmets can reduce the risk of certain injuries, such as skull fractures, whilst mouthguards have greatly diminished the severity of orofacial sports injuries (Daneshvar et al. 2011, Knapik et al. 2007). However, even with the use of such protective equipment, concussions - potentially debilitating head injuries - cannot yet be fully prevented. This is because concussions have a high chance of reoccurrence in those who have been concussed before



Figure 1.1: Due to the high-speed nature, collisions (both between players as well as between a player and a hard surface) occur often in hockey. To mitigate the damage done to the players when collisions occur, players wear a whole array of protective equipment, including helmets and mouthguards for hits to the head and face, body armour to lessen bruising to the body, and padding around major joints. (Image credit: Simon Emm, 2019)

(Daneshvar et al. 2011). Furthermore, conventional protective equipment does not help prevent overtraining and overuse injuries.

Contemporary sports safety equipment does not proactively inform the user about how they can best manage their health and keep safe. On the contrary, studies suggest that protective gears such as helmets may give players a false sense of safety, which can lead to riskier play-styles (a phenomenon often referred to as "risk compensation") which ultimately result in more severe injuries (Hagel & Meeuwisse 2004). The importance of the player's history in understanding the impact of injury, beyond any

isolated incident (Maron et al. 1998, Daneshvar et al. 2011), highlights the importance of protecting against prospective injuries. It also shows the value of having accurate data that can be used to optimize the management of an individual player. In light of this, I envision the future of protective equipment for sports to lie in solutions which more proactively engage with athletes’ health and well-being via instrumented sensors, and prevent injuries through data-driven and analytical solutions.

1.4 The State Of On-field Monitoring For Better Performance And Safety

Analytics in sports was first attempted in the 1950s by Charles Reep and Bernard Benjamin, who manually recorded tactical data and game states from professional soccer games. In recent decades, data analytics in sports have gained popularity thanks to advances made in computer vision and sensor technologies. Early sports analytics devices were made to track players’ movements, and help coaches analyze tactical performance post-game (Apostolou & Tjortjis 2019). More recently, player-tracking devices have been made to capture physical and physiological data from athletes unobtrusively in an on-field setting. Data from these player-tracking devices allow coaches to monitor the health and physical states of their players in real-time. This allows coaches to actively reduce the injury risk to players through giving much-needed breaks in training and substituting overexerted players in the game. Being able to capture the players’ physical demands in-game can allow coaches to create more game-relevant training routines. In practice, tracking players’ physical expenditure in practice can help avoid overtraining and protect the players from injuries and illnesses (Drew & Finch 2016). With all these exciting possibilities in mind, the International Football Association Board principally allowed the usage of wearable player tracking devices during games (Lutz et al. 2020), and many other sports soon

followed (Apostolou & Tjortjis 2019).

Currently, most on-field player monitoring techniques fall into one of two categories - workload monitoring and impact monitoring (Cheng & Bergmann 2022). Impact monitoring involves measuring sudden and forceful collisions — either between players or between a player and the playing surface — with the goal of detecting potential injuries. While some studies have investigated methods to capture body collisions for the purpose of musculoskeletal injury detection (Naughton et al. 2020, Gabbett et al. 2011), most developments in impact monitoring are concerned with measuring head impacts to detect traumatic brain injuries (TBI) (O’Connor et al. 2017). Yet, despite all the efforts, head impact measurement devices have failed to meaningfully improve the health and well-being of athletes due to their high error rates and low specificity in diagnosing concussions (O’Connor et al. 2017, Cheng & Bergmann 2022, Haste et al. 2025). Evidence suggests that head impact measurement devices are inherently flawed as they are only capable of measuring acceleration of the skull and not the forces experienced by the brain.

Workload monitoring, on the other hand, measures the physical work experienced by the athletes (external workloads), as well as the physiological and psychological responses to the physical work (internal workloads) (Bourdon et al. 2017). External workloads are usually captured via some form of positioning systems (e.g. Global Positioning System - GPS, Local Positioning System - LPS, Cameras, Inertia Measurement Units - IMUs) that measure the kinematics of the players (Benson et al. 2020). Understanding the external workloads experienced by players during games is extremely helpful in designing training routines that better prepare the players for games (Oliva-Lozano et al. 2021).

Internal workload measures are less straightforward than external workload measures, encompassing a broad range of physiological metrics (such as heart rate (HR),

breathing rate (BR), minute ventilation (VE), volume of oxygen uptake (VO₂), blood oxygen saturation (SpO₂), core body temperature, and more) as well as subjective psychophysiological measures like rating of perceived exertion (RPE) and well-being questionnaires. Outside of lab settings, RPE and HR are the two most commonly tracked internal workload metrics (Cheng & Bergmann 2022). Studies have found a correlation between perceived exertion and player wellness (Perri et al. 2021, Clemente et al. 2018), and that RPE can potentially be used to forecast the likelihood of injury when combined with external workload metrics (Vallance et al. 2020).

1.5 Respiratory Signals as Markers of Internal Workload

Several respiratory system related physiological measurements - such as breathing rate (BR; A.K.A. respiratory rate, RR; respiratory frequency, Rf), minute ventilation (VE), tidal volume (VT), and volume of oxygen uptake (VO₂) - are commonly used to evaluate an athlete's fitness level in lab-based exercise testing (Löllgen & Leyk 2018). One theory suggests that breathing techniques which couple breathing pattern with locomotion can improve endurance in running sports by reducing the workload placed on the respiratory muscles (Daley et al. 2013). Supporting this theory to a degree, controlled breathing technique trainings have been shown to improve the fitness and well-being of a population of chronic heart failure patients (Bernardi et al. 1998). Additionally, in the realm of performance monitoring, BR and VE have been shown to correlate closely with perceived exertion (de Almeida e Bueno et al. 2021, Nicolò, Montini, Girardi, Felici, Bazzucchi & Sacchetti 2020). And yet, despite this evidence supporting the importance of breathing monitoring in sports, it has yet to be realized in practice (Nicolò et al. 2017).

Within the wearable technology space, several methods to measure breathing rate

exist. Mainly, four categories of technology were observed (Massaroni et al. 2019, Liu et al. 2019, Costanzo et al. 2020, Vanegas et al. 2020, Vitazkova et al. 2024): modulation of other physiological signals, such as ECG (Charlton et al. 2017) or PPG (Chang et al. 2018, Charlton et al. 2017); chest wall volume changes and movements measured via chest straps or instrumented garments, such as Respiratory Inductive Plethysmography (Rahman et al. 2024), piezoresistive strain sensors, inertial sensors (De Fazio et al. 2021), etc.; measurement of air flow properties, such as shifts in temperature, moisture, chemical compounds (Kim et al. 2024, Yin et al. 2024), or direct measurement of respiratory airflow via a flowmeter (Massaroni et al. 2019); and lastly, acoustic sensing of respiratory sounds (Nam et al. 2015, Martin & Voix 2017, Harvill et al. 2022, Kumar et al. 2021, Romano et al. 2023, Lim et al. 2022).

Through a series of literature reviews (detailed in Chapter 2), it was determined that the microphone-based method has the greatest potential for contact-sport applications, particularly when factors such as device form factor, cost, usability, user comfort, and robustness to motion artifacts are considered. Thus, it was decided that this project would focus on the development of a microphone-based respiratory sensing platform, with the intention of capturing a range of respiratory metrics from breathing audio, and establishing evidence for the use of respiratory metrics in exertion monitoring.

1.6 Overview of Data Science Techniques

Various data manipulation and machine learning techniques were leveraged to aid in the development of a microphone-based wearable respiratory monitoring device. This section presents an overview of these techniques and their relevance to the project.

1.6.1 Audio Signal Processing

Audio signals are time series data by nature, though attempting to find patterns and interpret meaning from the raw data, especially in the time domain, can prove difficult. Feature engineering techniques have therefore been devised to help extract information from audio data. For starters, frequency filters are often applied to the signal, both to eliminate undesirable noise in the signal as well as to give emphasis to frequencies that contain the information most useful for the task at hand. For example, for breathing-related tasks, it’s been found that most information is contained within frequencies below 2kHz (Gavriely et al. 1981, Sarkar et al. 2015). The Butterworth filter is often the preferred type of frequency filter for this purpose (de Almeida e Bueno et al. 2021, Cheng & Bergmann 2023, Romano et al. 2023).

Amplitude Envelope Extraction

Amplitude envelope extraction is a fundamental technique in audio signal processing. It involves isolating the smooth, time-varying contour of a signal’s amplitude. It is used to extract information in the time domain, and can be used for temporal feature classification and temporal pattern detection. Common methods for amplitude envelope extraction include Hilbert Transform, windowed root-mean-squared values, and combining full-wave/half-wave rectification with either a moving average filter, a low-pass filter, or a peak detection algorithm. When applied to respiratory audio, the envelopes extracted are representative of the modulation wave emerged from breathing events, where each period corresponds to a breathing phase.

Frequency Domain Feature Engineering

To further unpack information from audio signals, frequency domain analyses are often utilized. Pure frequency domain analysis can be carried out using a Fourier Transform, which can be used to identify the dominant frequency in the signal. This

can be used for tasks such as measuring the beat tempo of music, or estimating a person’s breathing rate, granted the person is breathing at a steady pace.

When a piece of audio data is finely segmented and has a Fourier Transform applied to each segment, a data representation that is in both time and frequency domains is created, called a spectrogram. This technique is referred to as a Short-Time Fourier Transform (STFT). The spectrogram which results allows for the detection and identification of more complex patterns in the signal than raw time-series waveforms. Other types of spectrogram exist, with different strengths according to their different applications. Among these, Mel spectrograms are the most widely applied in audio signal processing at present, especially in deep-learning applications. This is partly due to Mel spectrograms being more computationally cost-effective than alternative audio representations (Purwins et al. 2019). Mel spectrograms are spectrograms with nonlinear frequency domain scaling that mimic the frequency scaling of the human auditory system. To create a Mel spectrogram, the magnitude spectra of a typical spectrogram, computed via STFT, are projected into a reduced set of frequency bands. In this project, Mel spectrograms have been used to identify breathing events, as well as to enable computer-vision models designed to estimate minute ventilation from breathing audio.

1.6.2 Machine Learning

Machine Learning (ML) is a data science field focused on developing statistical models that allow computers to learn patterns from data and make predictions or decisions without being explicitly hard-coded to perform specific tasks following rigid programming (Mitchell & Mitchell 1997). Machine learning can be broadly categorized into three categories: i) Supervised learning, where the model is fitted to training data that is labelled with a ground truth. Supervised learning techniques are often applied to classification and regression problems. ii) Unsupervised learning, where the model

is designed to spot patterns in unlabelled data. Unsupervised learning techniques are often applied to clustering and dimensionality reduction problems. iii) Reinforcement learning, where the model is given a basic set of legal actions and knowledge of its environment, and the model learns to perform tasks and navigate the environment through a reward system. Reinforcement learning is often utilized in automation, personalized recommendation systems, and game engines. In this work, we will be entirely focusing on classification and regression problems utilizing supervised learning techniques.

Linear Regression

Linear Regression is the most basic form of supervised machine learning. A linear regression model can be expressed as:

$$y = \vec{W} \cdot \vec{X} + b \quad (1.1)$$

where y is the output value, $\vec{W}$ is the weights, $\vec{X}$ is the input vector, and b is the bias. The weights and the bias are calculated to minimize the mean squared error between predicted values and the ground truth. In this thesis, linear regression was used to model exertion using cardio-respiratory features.

Decision Tree & Random Forest Regression

A decision tree is an algorithm that predicts outcomes by learning decision rules based on the input features, and branching into different values based on the decision made. Decision trees are computationally simple and easy to interpret, but are prone to overfitting. Random forest regression (RFR) is an ensemble method that builds multiple decision trees and combines their outputs to make more robust and accurate predictions (Breiman 2001). Each tree is trained on a random subset of the data,

and each split in the tree considers only a random subset of features, adding diversity among the trees. The final prediction is typically the average of the predictions from all individual trees. This aggregation reduces overfitting and improves generalization compared to a single decision tree, since the generalization error converges to a limit as the number of trees in the forest grows larger. In this thesis, RFR was used to model exertion using cardio-respiratory features.

Deep Learning

Deep learning is a subfield of machine learning where the algorithms are inspired by the structure of biological neural networks. Thus, deep learning algorithms are often referred to as deep (artificial) neural networks (DNNs). In contrast to linear regression and decision trees, DNNs are non-interpretable black boxes (Yedida & Saha 2021). DNNs are typically constructed from computational layers, referred to as hidden layers, sandwiched between the input and the output layers. The layers consist of computational units called neurons, and each hidden layer is connected to the next via activation functions (which can either be linear or non-linear functions). Non-linear activation functions often play a crucial role in allowing the DNNs to learn non-linear relationships between the inputs and the outputs.

DNNs typically achieve learning by adjusting the weights associated with each neuron in order to minimize the disagreement between the output and the ground truth. The mathematical representation of this disagreement is referred to as the loss function (or cost function, often denoted by L). The weights are adjusted through a process called gradient descent, in which partial derivatives of the loss function are computed in order to minimize the loss:

$$w \leftarrow w - \eta \frac{\partial L}{\partial w} \quad (1.2)$$

here, W is the weights and η is the learning rate. The partial derivatives of the weights from each layer are calculated backwards, starting from the output layer and ending with the input layer, using the chain rule:

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w} \quad (1.3)$$

where a is the activation function, and z is the computational output of the neurons in that layer. This method is referred to as backpropagation.

Multilayer Perceptron

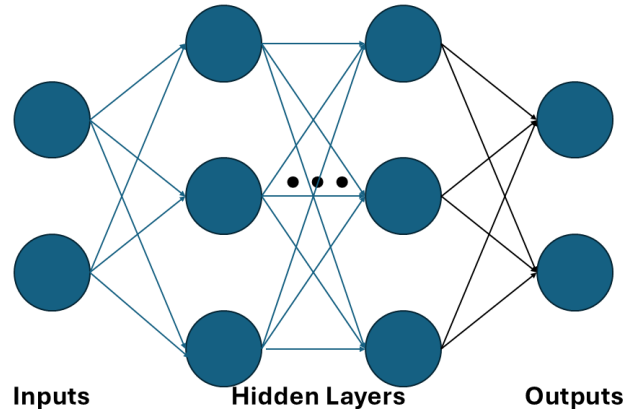


Figure 1.2: An illustration of a multilayer perceptron. In an MLP, every neuron is fully connected to the adjacent layers, meaning each neuron can be expressed as $f(b + \sum w_i x_i)$, where f is the activation function, the x terms are the outputs of the neurons from the previous layer, the w terms are the weights connecting those neurons to this one, and b is the bias.

Multilayer perceptrons (MLPs) are the most basic type of DNN. It has a simple feed-forward network consisting of multiple layers of neurons (see figure 1.2). MLPs are fully connected networks, meaning each neuron in one layer is connected to every

neuron in the next layer (Popescu et al. 2009). MLPs are commonly used for light-weight classification and regression tasks. In this thesis, MLPs were used to model exertion using cardio-respiratory features, and to predict minute ventilation from Mel spectrograms of breathing audio.

Convolutional Neural Network

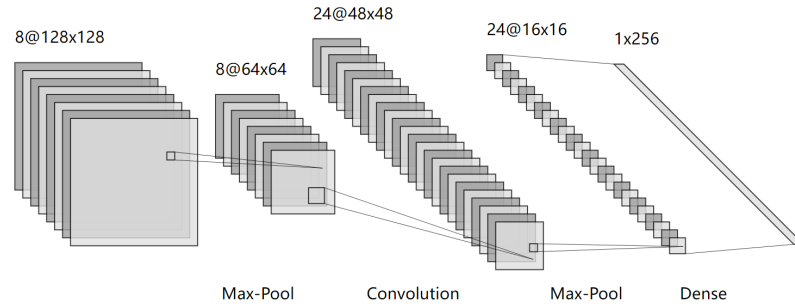


Figure 1.3: An illustration of a convolutional neural network. The convolutional layers apply filters (kernels) to local regions of the input to extract image features. These features are then passed through activation functions (e.g., ReLU) and reduced in size using pooling layers before they are passed onto the next convolutional layer. The network typically ends with one or more fully connected layers that combine the features to make predictions. (Image credit: Xinyu Hou (2023))

Convolutional Neural Networks (CNNs) are a type of DNN that are inspired by the biological visual perception system, and is purposefully designed to process images (Li et al. 2021). In a CNN model, filters called CNN kernels are applied to the data input, extracting features such as edges, corners, textures, etc. from the input image, mimicking biological receptors that can respond to various features (see figure 1.3). In audio processing, CNNs are typically used for sound classification using spectrograms; or as a part of a larger, more complex network, acting as spectrogram

feature extraction tools (Amodei et al. 2016). In this thesis, CNNs were used to extract features from Mel spectrograms of breathing audio, which are then used to predict minute ventilation.

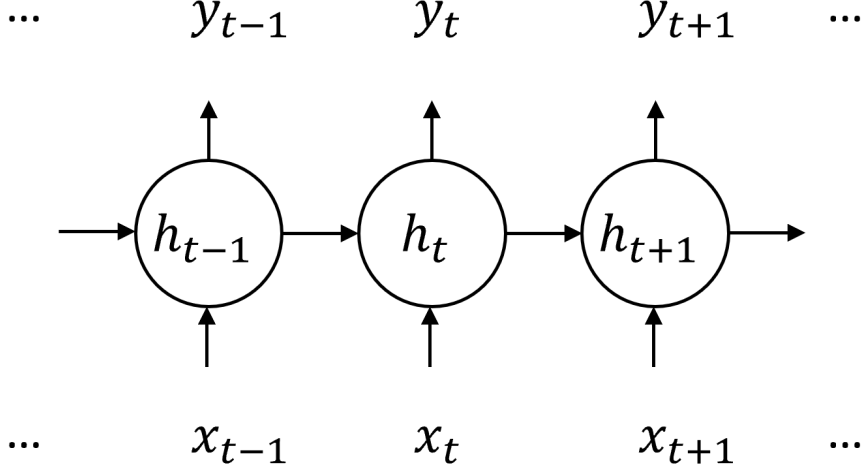


Figure 1.4: An illustration of a standard recurrent neural network. The output of the current timestep, y_t , is a function of both the input for the current timestep, x_t , as well as a hidden state computed at the previous timestep, $h_t - 1$. (Image credit: Xinyu Hou (2023))

Recurrent Neural Network

Recurrent neural networks (RNNs) are a type of DNN that are purposefully designed to handle sequential data (Elman 1990). RNNs compute information for each time step using information at that time step as well as a hidden state computed from the previous time step (see figure 1.4), and are therefore more capable of solving complex temporal problems than other types of DNNs. However, RNNs are prone to vanishing gradient issues during backpropagation, making them impractical for long data sequences. To mitigate the vanishing gradient problem, memory cells and gating mechanisms are introduced to maintain a more stable flow of gradients during backpropagation, forming a special RNN architecture named Long Short-Term Memory

(LSTM) (Hochreiter & Schmidhuber 1997). In audio processing problems, “future” context can often matter just as much as “past” context offered by standard RNNs, and thus, a special RNN variation called bi-directional RNN (Bi-RNN) is commonly used in audio analyses (Purwins et al. 2019). A Bi-RNN is effectively two RNNs processing the input data in parallel, one reads the input sequence forward in time, and the other reads it backwards, providing two hidden states computed from both the previous and the following timesteps (Schuster & Paliwal 1997). In this thesis, bi-directional LSTMs (Bi-LSTM) were used to predict minute ventilation from Mel spectrograms of breathing audio.

1.7 Chapter Outlines

The goal of this project is to develop novel on-field player monitoring methods that can potentially close gaps in current on-field monitoring practices, and promote data-driven approaches to player safeguarding in team sports. Through a literature review conducted at the beginning of this project (detailed in Chapter 2), a gap in physiological measurement in sports was identified. Respiratory information monitoring has been identified as a potential way to help gauge the on-field physical exertion of athletes and advise for better health and well-being safeguards. Following this discovery, we set out to develop algorithms along with an accompanying sensor system that can be used to capture and process respiratory data, and estimate physical exertion in sports settings.

The rest of the thesis is structured as follows:

- Chapter 2 details the exploration of the existing literature to identify what gaps exist in current on-field athlete-monitoring practices. The chapter begins with a systematic scoping review of on-field monitoring in team sports, which shows that current approaches are dominated by workload and impact monitor-

ing, and that internal workload monitoring relies mainly on heart-rate-derived measures and subjective ratings of perceived exertion. This finding identifies respiratory monitoring as a potentially valuable but underused area. The chapter then uses an umbrella review of wearable respiratory-monitoring technologies to examine how breathing can be measured outside the laboratory. This review compares approaches based on indirect physiological estimation, chest-wall movement, airflow properties, and acoustic sensing, and identifies acoustic sensing as a promising option for sport-specific use. Finally, a rapid review of microphone-based respiratory monitoring in sport and exercise is conducted. This final review shows that microphones can be used to estimate breathing during physical activity, but that existing systems are largely designed for controlled or non-contact sport settings. Together, these reviews establish the need for a microphone-based respiratory-monitoring system that is suitable for team contact sports.

- Chapter 3 addresses the question of whether breathing can be captured unobtrusively using a microphone integrated into sports protective equipment. Building on the gap identified in Chapter 2, this chapter compares possible sensor placements and evaluates whether an oral-cavity-based microphone can capture respiratory audio with sufficient quality for breathing-rate estimation. It describes the development of a microphone-instrumented mouthguard. The findings show that oral-cavity-based acoustic sensing can capture high-quality respiratory signals and support breathing-rate estimation under controlled conditions, while also revealing the need for more robust algorithms and a larger sport-specific dataset.
- Chapter 4 details a large-scale data collection which took place after the preliminary explorations documented in Chapter 3. This chapter presents a Yo-Yo Intermittent Recovery Level 1 testing protocol in which breathing audio

is recorded using the instrumented mouthguard alongside reference respiratory measurements, heart rate, GPS-derived running speed, and ratings of perceived exertion. This dataset provides the foundation for the algorithm-development and modelling studies presented in Chapters 5–7.

- Chapter 5 addresses the question of whether respiratory audio captured by the mouthguard can be used to detect individual breathing events and estimate breathing rate during high-intensity intermittent exercise. Using the dataset collected in Chapter 4, breathing audio is transformed into mel spectrograms, converted into a respiratory sound envelope, and analyzed using a trough-detection algorithm. The algorithm is evaluated against manual annotations and COSMED K5 breathing-rate measurements. The results show that mouthguard-recorded audio can support accurate breathing-event detection and practical breathing-rate estimation during outdoor exercise.
- Chapter 6 addresses the question of whether perceived exertion can be estimated from objective cardiorespiratory data, and which physiological measures are most informative for this task. Using the physiological and RPE data collected in Chapter 4, linear regression, random forest regression, and multilayer perceptron models are compared for estimating RPE. Feature-importance analysis shows that minute ventilation is the most important predictor of RPE, followed by heart rate. This finding establishes evidence that respiratory data can support perceived exertion estimation, and directly motivates the focus on estimating minute ventilation from breathing audio in Chapter 7.
- Chapter 7 addresses the question of whether minute ventilation, identified in Chapter 6 as a key predictor of perceived exertion, can be estimated directly from mouthguard-recorded breathing audio. Using the Chapter 4 dataset, this chapter evaluates deep-learning approaches for mapping respiratory audio to

COSMED K5 minute-ventilation measurements. The findings show that audio-based deep learning can extract meaningful information about minute ventilation from oral-cavity respiratory audio, supporting the broader thesis aim of using microphone-based respiratory monitoring to estimate exertion-relevant physiological information during sport.

- And finally, in Chapter 8, we reflect on all the discoveries made throughout this project, and explore some future directions this research could take.

Chapter 2

Literature Review

Chapter Abstract

This chapter presents three literature reviews undertaken to identify gaps in on-field sports monitoring for player safeguarding and to define the focus of this thesis. First, a systematic scoping review examines current practices for on-field physical and physiological data capture in team sports aimed at improving athletes' health and well-being. Three databases, Web of Science, PubMed, and IEEE Xplore were searched using keywords derived from the research question, and 2584 papers were screened to find 252 eligible papers to be included in the review. The review found that most monitoring approaches focus on workload tracking or physical impact monitoring. The review highlighted persistent issues with inconsistent metrics, limited standardization, and insufficiently large and diverse datasets. Internal workload monitoring methods are largely limited to survey-based ratings of perceived exertion and heart rate derived measures, and core temperature but only at the highest levels of play. Meanwhile, breathing has been identified as a clinically meaningful but under-utilised parameter in athlete monitoring.

To explore this gap, an umbrella review was conducted on wearable methods for

respiratory monitoring. The Web of Science database was searched for review articles on wearable technologies for respiratory monitoring. 49 review articles were retrieved and screened, of which 15 review articles were found eligible and included in the review. Four broad approaches were identified: indirect estimation from other physiological signals, chest wall movement sensing, airflow-property sensing, and acoustic sensing. Although wearable respiratory monitoring is advancing rapidly, most systems have been developed for healthcare, laboratory, or low-movement settings, with limited attention to the practical demands of team contact sports. Of the identified methods, acoustic sensing was deemed to be the most feasible for integration into contact sports gear.

Finally, a rapid review examined microphone-based respiratory monitoring in exercise and sport. Three databases, Web of Science, PubMed, and IEEE Xplore were searched using keywords derived from the research question, and 84 original research papers were screened to find 7 eligible papers to be included in the review. This review found that microphone-based systems are a feasible approach for measuring breathing during physical activity, with wearable implementations including headphones, in-ear devices, and sensors positioned near the mouth and nose. The literature shows that these systems can estimate respiratory rate with good accuracy under controlled conditions, typically using peak-detection or deep learning approaches applied to breathing audio. While demonstrating good potential for respiratory monitoring in non-contact sports such as running and cycling, current microphone-instrumented wearable form factors are incompatible with the demands of contact sports.

Overall, the reviews identified a need for a robust, unobtrusive, sport-specific wearable solution for monitoring breathing in team contact sports.

2.1 On-field Data Monitoring Techniques to Track Health and Well-being of Team-sports Athletes: A Systematic Scoping Review

2.1.1 Introduction

Participation in sports has become an essential part of healthy living in today's world. However, injuries can often occur during sports participation. With advancements in sensor technology and data analytics, many sports have turned to technology-aided, data-driven, on-field monitoring techniques to help prevent injuries and plan better player management. The purpose of this review study is to analyze the current practices of on-field data analytics in team-based sports which could help to improve the health and well-being of the athletes. This review will also highlight some of the novel techniques currently in development.

A version of this work has been published as "Impact and workload are dominating on-field data monitoring techniques to track health and well-being of team-sports athletes" in *Physiological Measurement*, but the work presented here has been modified to reflect the narrative of the thesis.

2.1.2 Methods

For this review, three databases, Web of Science, IEEE, and PubMed were searched using keywords connected by Boolean operators. The set of keywords used was: 1. (((('team' OR 'team-based') AND ('sport' OR 'sports')) OR ('football' OR 'soccer' OR 'Australian Rules' OR 'baseball' OR 'basketball' OR 'hockey' OR 'rugby')) 2. ('monitor' OR 'monitoring' OR 'track' OR 'tracking') 3. ('health' OR 'well-being' OR 'wellbeing' OR 'injury' OR 'impact' OR 'performance' OR 'exertion' OR 'overexertion' OR 'fatigue' OR 'recovery'), and the three fields are connected by the Boolean

operator AND. The keywords were searched in Topic Field for Web of Science, All Metadata for IEEE, and Title/Abstract for PubMed. All papers published before October 2021 were included in this search.

This review aims to investigate how physical and physiological data captured on-field can help improve athletes' health and wellbeing. Studies that use player-monitoring systems to investigate tactical performance, technical performance, or team dynamics are excluded. Papers that describe data capture frameworks without discussing their relevance to improving athletes' health or wellbeing are also excluded. In the search terms, we have specifically included the most popular competitive team-based sports with high revenue generation, based on reports from professional sports leagues, since they are the most likely to invest in or explore player-monitoring techniques, though, studies based on other team-based sports are not excluded if they are on-topic. Only papers published in English are included in this review. The retrieved papers were categorized. A time-series trend analysis was done, using a Mann-Kendall test, to analyze the growth of respective topics within the field.

2.1.3 Findings and discussions

The literature search found 2584 papers. After screening through the retrieved papers using the parameters defined above, 252 papers were found to fit the eligibility criteria. Two major categories of monitoring methods were identified, with 200 papers discussing workload tracking, 44 papers discussing impact monitoring, and 6 papers discussing both. Within impact monitoring, there is a strong focus on head impacts ($n = 40$). There were also 6 papers found that reported monitoring methods that do not fall under those two categories, 4 of which reported these methods alongside workload and impact monitoring. Such methods include biochemical sensors for monitoring hydration levels, COVID-19 exposure monitoring via positional data, etc. Figure 2.1 shows the key locations of monitoring devices found in current literature.

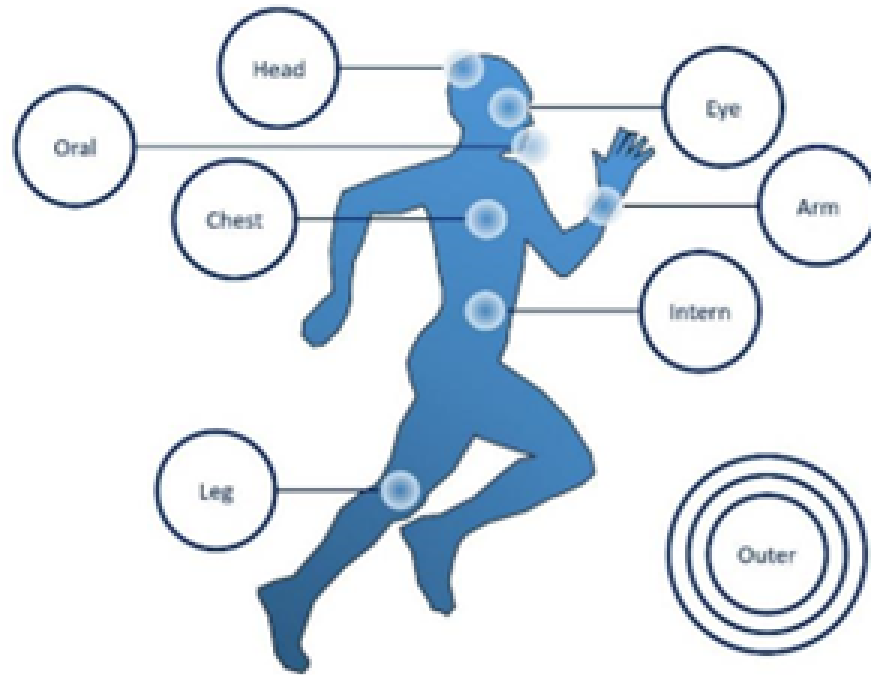


Figure 2.1: Types of monitoring devices based on placement. Head sensors are placed on e.g. helmets, headbands, caps or stick patches. Eye placement includes smart contact lenses, whilst sensing in the oral cavity is often done by an instrumented mouthguard. A shirt or strap is regularly used for sensing physiological or positional data at the chest level. Ingestible sensors can be used for internal monitoring (such as core temperature). The legs allow for a range of attachment possibilities including epidermal patches. Finally, off-body (outer) techniques can be applied for non-contact sensing for which multi-camera video analysis is one of the most widely used methods.

A Mann-Kendall Test on the number of studies published on workload monitoring for the past 10 years gives a Kendall Score of $S = 49$ and a two-sided p-value of $p = 0.00017$, strongly suggesting there has been an upward trend in workload related studies. A Mann-Kendall Test on the number of studies published on impact monitoring for the past 10 years ($S = 33$, $p = 0.011$), suggests that there has also been an upward trend in workload-related studies. It was found that workload monitoring

systems and measures have been adopted by major sports teams as a routine part of their training and matches, while head impact monitoring systems are still predominantly used as research devices. The following subsections will give a comprehensive synopsis of the findings.

2.1.4 Workload monitoring

Sports participation has become an essential part of ensuring healthy living in today's world. Team sports are organized at a variety of skill levels and are played by people in all age groups. Sports participation can bring tremendous benefits to both physical and mental health (Warburton et al. 2006). However, there are risks that come with participating in sports. Athletes can succumb to musculoskeletal injuries due to overuse of particular body parts (DiFiori et al. 2014). Muscle injuries can occur when the stresses and strains applied to body tissue exceed the body tissues' maximum capacities (Benson et al. 2020). In addition to physical injuries, an athlete who is over-reaching or overtraining can experience a drop in physical performance which requires time to recover from, and can sometimes suffer from negative physiological symptoms (DiFiori et al. 2014). One way of mitigating all these risks is by monitoring the workloads endured by the player.

Athlete workloads can be divided into two types, internal and external loads. External loads are quantifications of physical tasks performed by an athlete and are measured in objective movement-based metrics (Benson et al. 2020). Internal loads, on the other hand, are physiological and psychological responses an athlete has towards the external workload and incorporate both objective physiological measures and subjective measures of perceived exertion (Bourdon et al. 2017). Table 2.1 shows a list of technology used to monitor internal and external loading.

Table 2.1: A list of methods used to measure internal and external loading in the literature.

Technology	Mechanism	Measures
External load		
Global positioning system (GPS)	Signal transmissions from multiple orbital satellites and a ground-based receiver. The relative delay in the signal is used to calculate the position and speed of the receiver.	Positioning Data: Distance, Velocity, Acceleration, Jerk; Derived Measures: PlayerLoad, Metabolic Power, Number of Jumps, Number of Turning Events, etc
Local positioning system (LPS)	Signal transmissions from multiple local antennas and a ground-based receiver. The relative delay in the signal is used to calculate the position and speed of the receiver.	
Cameras	Multiple cameras placed at different angles around the stadium; footages are used to track player movements.	

Accelerometer/inertial measurement unit (IMU) that are body-worn*	A mass attached by a spring moves relative to two electrodes under acceleration, changing the capacitance.	* Pitch Count (for wrist-worn technology)
Internal load		
Subjective measures	Self-report surveys	Rating of perceived exertion (RPE), session RPE, visual analogue scale, likert scale, hooper's index, etc
Electrocardiogram (ECG)	Electrodes placed in a chest strap to measure electrical activity from the heart.	Heart rate, heart rate variability, training impulse, etc
Photoplethysmography (PPG)	Optical sensor place over blood vessels to measure heart rate via light absorption and reflection by the blood.	
Surface electromyography (EMG)	Electrodes placed on the skin to measure electrical activity from muscles.	Muscle fatigue
Ingestible temperature sensor	Ingestible capsule with a temperature sensor and a transmission circuit.	Core temperature

Epidermal patch/Biochemical sensing mouthguard	Biochemical sensors that can measure chemical concentrations in sweat or saliva.	Lactate and glucose concentrations as indicators for muscle fatigue Concentration of electrolytes can be used to measure hydration level
Near-infrared spectroscopy	Optical sensor that uses the Beer-Lambert law of light attenuation.	Muscle oxygen saturation

It was found that many workload monitoring methods can apply across a wide range of different sports, and that most methods can be grouped into one of two sports groups, contact sports and throwing sports. The differences in play styles between these two groups of sports have resulted in differently defined workload measures that require the utilization of different monitoring techniques.

2.1.4.1 External workload monitoring in contact sports

External loads are often determined by positioning and accelerometry-based metrics in contact sports. By capturing the players' in-game external load demands and monitoring their training (external) loads, sports trainers can optimize training sessions to better prepare their players for games (Oliva-Lozano et al. 2021). Video tracking via multi-camera systems was used to monitor players' movements on the pitch and it was the most common player tracking method in soccer until 2014 (Castellano et al. 2014, Rico-González et al. 2020). Due to the high cost and installation difficulties, multi-camera systems were only installed in official match stadia and were mostly used for match assessments (Rico-González et al. 2020). To allow sports teams that

do not have access to multi-camera positioning systems to track their players' movements both in games and training, radio frequency-based (RF) tracking systems such as global positioning systems (GPS) and local positioning systems (LPS) have been adapted for sports purposes (Rico-González et al. 2020, Theodoropoulos et al. 2020). A GPS device can detect the location of its user through signals transmitted by four or more satellites and can determine the user's velocity via Doppler shift. GPS devices require reliable connections to satellites to function, so when a sports facility does not have sufficient satellite coverage, such is the case for many indoor games (Benson et al. 2020), LPS devices can be used in their place. LPS works on similar principles as GPS, with the satellites being replaced by locally installed antennas. LPS systems are claimed to have pinpoint accuracy, though, at present, their use has been limited due to the high cost of installation and calibration (Theodoropoulos et al. 2020).

Common sports GPS devices in the present-day sample at 10 Hz, though some can go as high as 18 Hz. While GPS devices can capture the position and velocity of players, they are less useful for observing rapid acceleration and change in direction due to their low refresh rate and their inability to determine orientation. To address this, many GPS-based sports monitoring systems are integrated with inertial sensor units (IMU) that consist of accelerometers, gyroscopes, and magnetometers. Most IMUs used in workload monitoring can record at 100–120 Hz (Theodoropoulos et al. 2020). Today, the most popular implementation of this technology is an integrated system that consists of a GPS, an IMU, and a heart rate (HR) monitor. An example of this is the Catapult Vector (Catapult Innovations, Scoresby, Australia) which has an up to 18 Hz GPS, a 100 Hz IMU, an electrocardiogram (ECG) derived HR monitor, and an ultra-wideband LPS receiver.

There is a common set of external load variables derived from the position-tracking systems mentioned above. The basics are the total distance (TD), velocity, and acceleration (Bourdon et al. 2017, Theodoropoulos et al. 2020). There are a large

number of external load measures derived from these base measurements, often as a data zoning of some sort, or a new measure calculated from the base measures using an algorithm. The frequencies of certain physical actions are derived from these measures and added up as external load measures, such as the frequency of turns and the number of jumps (Benson et al. 2020). Jerk may be derived from the acceleration and used as an external load measure as is seen in the case of PlayerLoad by Catapult (Boyd et al. 2011).

A commonly used external load measure calculated from the basic measures is the metabolic power. The metabolic power is an energy exertion estimation obtained from acceleration by establishing an equivalence between accelerating on a plane surface with walking uphill at a constant speed (di Prampero et al. 2005). The metabolic power does not account for interpersonal variability. In addition, the 'metabolic power' does not account for the energy used to overcome friction and air resistance at a constant speed.

There exists a large number of external load zoning metrics, such as high-speed distance covered (zoning of speed), sprint-running distance (zoning of either speed or acceleration depending on the study), acceleration intensity (zoning of acceleration), and so on (Benson et al. 2020, Rago et al. 2020, Theodoropoulos et al. 2020). Most papers covering these metrics have their own definition of terms such as 'high-speed' or 'sprint' and have their own ideas of how to set the thresholds for each zone (Benson et al. 2020, Kupperman & Hertel 2020, Rago et al. 2020, Theodoropoulos et al. 2020). One review found that the minimum threshold for 'high-speed' running in the literature varies from 14.4 to 24 km/h, and the minimum threshold for 'sprinting' varies from 19.8 to 25.3 km/h across different studies (Kupperman & Hertel 2020). These discrepancies make comparisons between different external load studies nearly impossible.

2.1.4.2 Internal workload monitoring in contact sports

Internal loads, in contrast to external loads, are not always captured in objective terms. One of the most frequently used methods to capture internal loads is Borg's scale of perceived exertion (RPE). RPE is used to quantify the conscious sensation of exertion (Marcora 2010). There are a few variations of RPE in the literature today, though they are all based on the same principle: subjective numerical measures based on the surveyed responses of the monitored athlete. The RPE has numerical values corresponding to verbal descriptors at some of the points on the scale. One way the RPE variations can differ from one another is the scaling of the numerical values corresponding to different verbal feedbacks. Another difference between some of the RPE variations is the time scale over which the RPE is captured. For example, the session RPE (sRPE), which is commonly used in conjunction with GPS systems in studies that aim to capture sports demands, is a one-off measure taken at the end of a sports session. There are a few other subjective internal load measures like the RPE, which aim to capture the perceived exertion of the athlete. Such measures include the visual analogue scale (VAS), which is a linear scale with two anchor points at the two extremes; and Likert scale, an RPE-like numerical scale with numerical descriptors at every point (Grant et al. 1999). Some survey-based internal load measures also aim to include the athletes' conditions prior to the match. One example of this is the Hooper's Index, which includes self-perceived ratings of fatigue, stress, delayed onset muscle soreness, and sleep (Haddad et al. 2013).

Whilst these subjective measures are often linked to the level of physical effort, they are also sensitive to stressors beyond physical effort such as psychological and environmental factors. And the surveying aspect, in practice, often requires a series of training and familiarization sessions between the athletes and dedicated coaches to produce consistent outputs, making them impractical to use at many levels of play (de Almeida e Bueno et al. 2021). The surveying aspect also means that these

measures are limited to non-realtime use, and cannot perform continuous monitoring throughout a session. All these limitations suggest that while subjective internal load measures provide insightful information on the athletes, they are not suitable as an on-field monitoring method on a session basis, but rather as something tracked over a long period.

On the objective end of internal loading measurement, the rate of oxygen consumption (VO_2) has long been used as a measure of performance and exertion in sports science. However, to directly measure VO_2 requires the use of obtrusive metabolic devices unfit for on-field monitoring. Heart rate (HR) is often used as an on-field measure for evaluating the intensity of exercise and has been shown to have a linear relationship with VO_2 over a large range of submaximal intensities (Li et al. 2016). Researchers have developed a series of HR-derived measures to evaluate internal loads, these include the ratio between the actual HR and the maximum HR, the ratio between the actual HR and the resting HR, binning of the ratios into different intensity zones, heart rate variability (HRV), and training impulse (TRIMP) which is computed from heart rate and duration (Stagno et al. 2007, Bourdon et al. 2017, Colosio et al. 2020). There are numerous studies on the abilities of HR-related metrics to capture fatigue and performance in various sports, though the actual effectiveness of these measures has been brought into question. Contradictory findings can be found in the literature and a better understanding of the science behind HR is pointing towards these HR-based measures being potentially a useful part of a multidimensional internal load monitoring system rather than as robust measures of performance, fatigue, and well-being in themselves (Schneider et al. 2018).

Alternative measures of internal loads have been investigated using different types of physiological sensors. Near-infrared spectroscopy (NIRS) muscle oximeters can measure muscle oxygen saturation (SmO_2) using the Beer-Lambert law of light attenuation. Vasquez-Bonilla et al. (2021) tested the use of a portable NIRS system

with women soccer players in small-sided games, and found that SmO_2 can be a useful measure for evaluating fatigue in players. Biochemical sensors that can detect electrolytes in sweat, and lactate and glucose concentrations in saliva and sweat have been built into integrated mouthguards and epidermal patches (Seshadri et al. 2016, Seshadri, Magliato, Voos & Drummond 2019, Seshadri, Li, Voos, Rowbottom, Alfes, Zorman & Drummond 2019). Lactate and glucose concentrations can be used as indicators of fatigue level, and electrolytes can be used to monitor hydration (Seshadri, Li, Voos, Rowbottom, Alfes, Zorman & Drummond 2019). A practice that has gained traction at professional levels in recent years is core body temperature sensing in the form of an ingestible pill (Muniz-Pardos et al. 2021). While this practice is effective, a less invasive and affordable alternative is desired at lower levels of play.

Although VO_2 and other respiratory metrics — such as breathing rate (BR), minute ventilation (VE), and tidal volume (VT) — are commonly used in lab-based sports performance testing (Löllgen & Leyk 2018), they have received relatively little attention in on-field monitoring (Nicolò et al. 2017). And while, VO_2 cannot be measured without analyzing the chemical composition of exhaled air, wearable solution to measure BR do exist in the literature (some common methods include photoplethysmogram (PPG), electrocardiogram (ECG), plethysmography, and breathing audio analysis) (Harbour et al. 2021, Martin & Voix 2017, Charlton et al. 2017). Studies have found that both BR and VE are closely linked with perceived exertion (de Almeida e Bueno et al. 2021, Nicolò, Montini, Girardi, Felici, Bazzucchi & Sacchetti 2020), and BR in particular has been found to have a faster response to recovery than HR. This could be the result of the psychophysiological nature of breathing (Wientjes 1992). Controlled breathing techniques have been shown to improve the fitness and well-being of chronic heart failure patients (Bernardi et al. 1998). It has also been hypothesized that breathing techniques that couple breathing pattern with locomotion can improve endurance in running sports by reducing the workload placed

on the respiratory muscles (Daley et al. 2013). By enabling the measurement of BR and VE accurately during outdoor exercises, we can put this theory to the test; and any potentially beneficial findings can be put into practice effectively.

Through this review, we found a wide array of physiological sensors that can offer realtime continuous on-field monitoring of players' internal loads in objective ways, which will allow for more timely intervention that will prevent athletes from harm. Moving forward, internal loads monitoring should rely on a wider array of physiological measures. In addition, future investigations should put a focus on how these measures relate to an athlete's likelihood of injury and long-term health.

2.1.4.3 Workload monitoring in throwing sports

Throwing athletes, such as baseball and softball pitchers, have a completely different set of physical demands compared to contact team sports athletes. As such, throw athletes require different workload measures and monitoring systems. Overuse injuries, often in the shoulder and elbow of the throwing arm, constitute more than half of the injuries experienced by baseball pitchers (Dowling et al. 2020). Pitch count is one of the conventional external load measures for baseball pitchers. Higher pitch counts have been linked to higher chances of experiencing pain or injuries (Dowling et al. 2020). Baseball guidelines have been introduced for adolescent players to restrict pitch counts in hope of reducing overuse injuries in pitchers, though implementing such guidelines requires coaches and players to have the awareness and resources to monitor the players. Schweiger et al. (2020) developed a wristband with a built-in IMU sensor that can detect and count throwing and pitching events. They demonstrated excellent accuracy in their initial test, successfully identifying all 161 throwing events in their test set with only one false positive. Another way to monitor workloads in pitchers is via surface electromyography (EMG) (Hettiarachchi et al. 2019). EMG electrodes are not obtrusive to gameplay in baseball as it is a non-contact sport, and

can thus be used on-field to monitor players' muscle fatigue. In addition to overuse injuries, pitchers can often injure themselves due to incorrect techniques (Stretch 1995). Along with pitch count and muscle fatigue monitoring, IMUs and surface EMGs can also be used to objectively screen a player's throwing technique (Hettiarachchi et al. 2019). Surface EMGs can be used to monitor the muscle activation pattern of the relevant muscles during a throw (Hettiarachchi et al. 2019). This can provide insights into a person's throwing technique. Through observation of low and high-risk muscle activation patterns, potential risk factors can be identified (Hettiarachchi et al. 2019). Another measurement that can provide a great deal of information regarding a person's throwing technique (and potential risk factors) is the joint angle kinematics, which can be measured using a set of appropriately placed IMUs. These sensors have the potential to improve injury prevention in throwing sports, especially in training for players who have yet to develop any throwing skills.

2.1.4.4 Workload measures and injury rate predictions

One of the most popular injury prediction measures currently used in sports science is the acute:chronic workload ratio (ACWR) (Kupperman & Hertel 2020). The acute workload is a measure of the current workload and is typically averaged over a week; the chronic workload is a measure of the workload the athlete has been prepared for and is typically averaged over 3–6 weeks. Time windows for both acute workloads and chronic workloads can sometimes vary between different studies (Kupperman & Hertel 2020). The ACWR is based on the theory that the acute workload is analogous to a state of fatigue and the chronic workload is analogous to a state of fitness, and when combining the two into a ratio, an index of athlete preparedness is obtained (Murray et al. 2017, Kupperman & Hertel 2020). Studies have established correlations between this preparedness index and injury risks, with a higher ACWR value indicating a higher risk of injury (Murray et al. 2017, Kupperman & Hertel

2020). In theory, the ACWR can be calculated for any workload measure, either internal or external. Both internal and external load ACWRs have been shown to positively correlate to injury risks (Murray et al. 2017). The average workloads in ACWR are either calculated as rolling averages (RA) that weigh workloads on all the days equally or as exponentially weighted moving averages (EWMA) that weigh the workloads experienced on more recent days greater than days further back in the past (Murray et al. 2017, Kupperman & Hertel 2020). It's been found that EWMA-based ACWRs are more sensitive than RA-based ACWRs for detecting changes in injury risks (Murray et al. 2017). The validity of using ACWR to predict the risk of injury is highly debated within sports science. One concern is that the popularity of ACWR is being propagated by editorials and commentary articles rather than research papers with quantitative analyses and that there currently is not enough evidence to validate its use (Kupperman & Hertel 2020). Another criticism against ACWR is that the acute workload is also a part of the chronic load, which might create a correlation between the two irrespective of any biological or physiological associations (Thornton et al. 2019). Despite its popularity, there seems to be a pushback against ACWR in recent years, with some recent studies showing strong evidence to reject any correlation between ACWR and injury risk (West et al. 2021).

Internal load measures and external load measures can be combined to paint a better picture of a player's physical state. The ratio between a player's external load and internal load during a training session can be used as a measure of the player's workload efficiency, which can indicate the player's fitness level (Grünbichler et al. 2020). Internal loads and external loads can also be combined to train a machine learning model that is capable of predicting injury risks. Vallance et al. (2020) trained and compared various injury prediction machine learning models based on internal and external load measures, and found models that use a combined input of internal and external loads performed better than models that use either internal or external

loads alone.

Currently, there is no consensus on how workload should be determined. There exists a large number of variants of the base metrics as well as numerous ways different workload measures can be combined into hybrid measures; one study reports up to 756 different combinations (Benson et al. 2020). No clear conclusion can be drawn on how to utilize workload monitoring for injury prevention, due to the heterogeneity of analytic approaches in the literature, and this issue is worsened by the fact that not all studies have the same definition of what constitutes an injury (Benson et al. 2020). There is an overwhelmingly strong need for standardization in the field of workload monitoring.

Nonetheless, there is little doubt that workload monitoring will become a standard practice in organized sports in the future. Monitoring players' workloads can potentially benefit the players' health as well as their growth as athletes. Such potential benefits include more optimized training based on game demands, lowered risk of overtraining, injury prevention, and smoother return-to-play transitions after injuries in the future. However, most studies on workload monitoring are based on statistically small data sets and are disproportionately skewed towards high-level male athletes (Benson et al. 2020, Kupperman & Hertel 2020, Rago et al. 2020). There is a need to collect data from a wider range of demographics. One of the potential limitations to data collection right now is the existing monitor systems are only affordable to well-funded sports entities. Future research should also include more developments of devices capable of measuring both internal and external loads, especially affordable devices that will be accessible at all levels of play.

2.1.5 Impact monitoring

2.1.5.1 Miscellaneous collisions monitoring

Physical collisions between players as well as impacts between a player and the play surface are a common occurrence in full-contact sports such as rugby football, Australian rules football, American football, and ice hockey. Collisions can often take place during tackling and checking actions which are integral to these games. These collisions can lead to soreness and muscle damage which may result in attenuated neuromuscular performance and recovery (Naughton et al. 2018). Tackling can also lead to injuries that cause players to miss training and matches (Gabbett et al. 2011). Monitoring collisions that contact sports players have to endure may help to mitigate the risk of injury. As with in-game workload monitoring, learning more about in-game collision loads can aid coaches in designing more specialized training and recovery strategies to better prepare their players (Wellman et al. 2017).

Collision loads are often quantified by the frequency and intensity of collisions (MacLeod et al. 2018, Naughton et al. 2020). In the past, collision load monitoring was done via video analysis, the number of collisions would be manually counted, and a perceived measure of collision intensity would be given following a quantitative analysis of the footage (Naughton et al. 2020). In more recent years, researchers have attempted to quantify collision intensity more objectively by extracting collision velocity and acceleration from videos frame-by-frame (Naughton et al. 2020). These methods are time-consuming and can be prone to human errors and biases.

GPS systems for external load tracking have become widely accepted by most football-code sports at a professional level. As stated previously, there are commercial systems that combine GPS with IMUs, which often include a 100 Hz tri-axial accelerometer, a gyroscope, and a magnetometer. Due to the popularity of these systems, many researchers have been attracted to investigate the viability of tracking

collision load using IMUs (Gabbett et al. 2011, 2014, Hulin et al. 2017, MacLeod et al. 2018). Attempts were made to identify collision events from the kinematic data captured by IMUs, though there is only limited validation of such systems in the literature (Naughton et al. 2020). One validated collision detection algorithm found in the literature works by thresholding the jerk-based PlayerLoad external load measure and detecting changes in orientation of the IMU (Gabbett et al. 2011, 2014, Hulin et al. 2017, Naughton et al. 2020). A collision is flagged when the PlayerLoad spikes above 2 AU (Arbitrary Units) following a change in orientation of the IMU (Hulin et al. 2017). This algorithm performs poorly when short-duration collisions (≤ 1 s) and low-intensity collisions (≤ 1 AU) are included (Hulin et al. 2017). When short-duration collisions and low-intensity collisions are excluded, the algorithm achieves a sensitivity of 93.9% and a specificity of 91.7% (Hulin et al. 2017). The inability of this algorithm to accurately detect so-called short-duration collisions and low-intensity collisions is an innate problem with thresholding. Perhaps, in the future, collision event detections can be made more accurate using more dynamic algorithmic approaches.

When monitoring collision events using accelerometers, the collision intensity is determined by the acceleration of the player due to the collision (Gabbett et al. 2014, Wellman et al. 2017, Naughton et al. 2020). In most studies, different ranges of acceleration are categorized into three severity levels, mild, moderate, and heavy (Gabbett et al. 2014, Wellman et al. 2017, Naughton et al. 2020). However, the acceleration ranges used for each severity level vary across different studies, which makes it difficult to meaningfully compare their outputs (Naughton et al. 2020). Presently, there's a clear need to standardize how collision intensity is quantified. In addition, current collision monitoring methods do not discriminate between different types of collisions: different types of collisions could have different implications on the athlete. Future work in collision monitoring should include applying more sophisticated methods, such as machine learning, to differentiate types of collisions (Naughton et al. 2020).

2.1.5.2 Head impact monitoring

Shocks to the head are often of special interest when it comes to monitoring body impacts due to their potential medical implications. Head impacts are often monitored with dedicated devices, separate from those measuring general impacts. Whether it occurs through direct collisions between athletes in tackling and checking, contact with the playing surface, or because of a blow to the head by a ball, traumatic brain injuries (TBIs) due to head impacts are of great concern as they can compromise the athletes' short and long-term health.

An estimated 300 000 sport-related concussions (mild TBIs) occur annually in the US alone, making up for 8.9% of all high school athletic injuries and 5.8% of all collegiate athletic injuries (Gessel et al. 2007). Some studies suggest that concussions only make up for between 8% to 19.2% of sports-related TBIs, making the number of US annual sports-related TBIs somewhere between 1.6 million and 3.8 million (Langlois et al. 2006). On-field diagnoses of TBIs often rely on self-reports from the affected athletes. This can sometimes lead to underreporting of TBIs, potentially due to transient or delayed symptoms, or players' unwillingness to leave play (McCrea et al. 2004, Williamson & Goodman 2006, LaRoche et al. 2016). Despite the recent evidence that has shown that policy changes can help reduce underreported incidents (LaRoche et al. 2016), an objective way to detect injurious head impacts in real-time is still warranted.

Head impacts are often analyzed as two separate mechanisms, translational (linear) movements and rotational (angular) movements. Both large linear accelerations (LA) and angular accelerations (AA) of the head can lead to injurious strains on the brain, with the latter associated with a higher risk of injury (Bailes & Cantu 2001). Due to the nature of TBIs, human experiments with deliberate injurious head impacts cannot be carried out, making acquiring data for TBI studies difficult. Human cadavers and

medical phantoms have been used in place of a live human in many past studies to grant insight on TBIs, though, how representative these substitutes are of the biomechanics of live humans remains an open question (Hardy et al. 2007, McAllister et al. 2012, Yang et al. 2006). And as of now, many of the clinical implications associated with head impacts sustained by athletes remain unknown (Bartsch et al. 2020, Mihalik et al. 2020).

In recent years, sensor systems have been developed to capture head kinematics of contact-sport athletes on-field to help bridge the gap between head impact data and TBIs, while simultaneously offering solutions to help flag TBIs in athletes in real-time. Development of such systems first started in American football, taking the forms of accelerometer arrays, such as the Head Impact Telemetry (HIT) System (Simbex, Lebanon, NH, USA), or inertial sensor units (IMU), such as the gForce Tracker (GFT) (Artaflex Inc., Markham, Ontario, Canada), which could be integrated into American football helmets (Campbell et al. 2016, Siegmund et al. 2016, Brennan et al. 2017, O'Connor et al. 2017). These helmet-mounted sensor systems can measure kinematics in six degrees of freedom, three linear and three rotational, and can estimate the location of head impact based on the kinematics. Raw kinematic data from these systems are often put through a correction algorithm to output the LA and AA at the center of mass of the head (Campbell et al. 2016). These techniques have also been adopted by other sports that include helmets in their kits, such as ice hockey and lacrosse. It has been reported that the accuracy of these systems is heavily influenced by the shape of the helmets they are in (Siegmund et al. 2016, Buice et al. 2018). Many limitations of these systems have been identified, with one of the greatest limitations being that helmets can move relative to the players' heads, especially when under impact. The helmet-mounted systems cannot account for the relative movements between the helmet and the head, which can lead to overestimated acceleration values when the helmet moves more than the head (O'Connor et al. 2017). The HIT system,

in particular, has also been critiqued for its high price and limited compatibility with different helmet models (Merrell et al. 2017). Furthermore, these systems can only be deployed in sports that use helmets, but TBIs are present in many sports that do not use helmets.

To accommodate for more sports, nonhelmeted systems with IMUs have been developed in the form of headbands, skullcaps, mouthguards, or xPatch (X2 Biosystems, Seattle, WA, USA)—a stick-patch that goes behind the ear (Brennan et al. 2017, Cummiskey et al. 2017, Merrell et al. 2017, O’Connor et al. 2017, Bartsch et al. 2020), with the latter two also addressing the issue of relative motions between the sensors and the player’s head. Like their helmeted counterparts, these nonhelmeted systems are designed to measure head kinematics and impact locations. These systems come with their own limitations. The accuracy of the stick patches can be affected by skin motions. And data transmission from the mouthguards can be affected by saliva accumulation (O’Connor et al. 2017).

In validation studies for these head-kinematics-based impact-monitor-systems, a Hybrid III (General Motors, Detroit, Michigan, USA) headform instrumented with sensors is often used as a reference for the measured kinematics (Duma et al. 2005, Jadischke et al. 2013, Campbell et al. 2016, Siegmund et al. 2016, Cummiskey et al. 2017, Merrell et al. 2017, O’Connor et al. 2017, Bartsch et al. 2020, Buice et al. 2018). The design of validation studies can sometimes be over-idealized and not representative of real-world usages. For instance, press-fitting a helmeted system onto the headform and effectively making the helmet and headform move as a single rigid unit (Duma et al. 2005, Cummiskey et al. 2017). Validation of the same device under more realistic conditions found much higher errors in the measurements (Duma et al. 2005, Jadischke et al. 2013, Cummiskey et al. 2017). Cummiskey et al. (2017) compared the accuracy of 2 commercially available helmet-mounted systems and 3 commercially available head-mounted systems and found that the head-mounted sys-

tems consistently outperformed helmet-mounted systems under lab settings. They suggested this was due to relative motions between the helmet and the headform, and since relative motions between the helmet and the player’s head are present in real life, this finding cannot be dismissed. Thus, Cummiskey et al concluded that to most accurately capture head acceleration events, it is necessary to mount the sensor system directly to the head. Cummiskey et al further suggested that under real-world conditions, the accuracy of the head-mounted systems such as the xPatch can be affected by skin motion, and therefore, moving forward, mouthguard-based sensors may be the most promising.

In terms of the actual measurement accuracies of these systems, the measurement quality can differ for different impact locations. Furthermore, inconsistent measurement accuracies of identical systems have been reported across different peer-reviewed works. The HIT System, for example, when compared with the Hybrid III, has been reported to have a relative error rate ranging from less than 4% to greater than 15% for LA measurements, according to studies from different researchers, and some have even observed peak linear acceleration measurements with root mean squared errors (RMSE) greater than 100% (Jadischke et al. 2013, Cummiskey et al. 2017, O’Connor et al. 2017). Based on this observation, a meta-analysis is likely needed to fully validate the accuracy of these systems, and till then, research teams intending to use these devices for data capture should perhaps carry out their own validation study whenever possible. Most IMU-based devices cannot measure angular acceleration directly but measure angular velocity via gyroscopes instead, and the angular acceleration is often numerically obtained via differentiating the angular velocity (Cummiskey et al. 2017, Buice et al. 2018, Rich et al. 2019). Head acceleration events often happen on a scale of milliseconds, as such, a minimum sampling rate of 1 kHz is desired, and higher sampling rates have been shown to further improve measurement accuracy (Cummiskey et al. 2017, Rich et al. 2019). Impact locations are estimated based

on the LA and AA measurements, and the outputs are divided into discrete impact zones. There's a fundamental flaw in the way the impact location is calculated, as both the impact location and the force direction are unknown, and it is, therefore, impossible to obtain a unique impact location using just the head kinematics (Cumiskey et al. 2017). This issue is slightly alleviated by the fact that the human head is not isotropic which makes some impact locations more likely than others for different head kinematics, but the issue is still ever-present. Helmets can be instrumented with tactile sensors to better measure the location of impact (Merrell et al. 2017), though there is no equivalent solution for non-helmeted systems such as mouthguards.

In certain sports, such as soccer, players are often reluctant to adopt additional accessories that are not already a part of their kit. To address this issue in soccer, Stone et al. (2018) proposed the development of a smart soccer ball with integrated IMUs for heading detections. In their study, the smart ball showed promise for heading detections under controlled testing conditions, though, Stone et al stated that better sensors are needed to make their system viable in the real world.

For the reasons listed above and more, the aforementioned sensor systems are not currently widespread amongst athletes outside of research studies. At a professional level, many sports have already adopted monitoring systems for capturing internal and external loads, and some researchers have proposed to detect and measure head impacts using data captured by some of these systems.

In soccer, it is already common for professional players to wear GPS vests with a built-in IMU located above the third thoracic vertebra (T3) of the spine. Worsey et al. (2020) investigated the potential to use that T3 IMU as a detection device for heading events in soccer. In their study, they found that the T3 IMU can only make reliable heading detections in athletes when certain techniques are used, and it cannot make any accurate head kinematics measurements, making it, unfortunately,

not suitable for heading detection and monitoring.

In American football, while the aforementioned head impact monitoring systems are not commonly employed by professional athletes, professional stadiums are equipped with high-quality camera systems that can capture game footage which will allow for analyses of impacts sustained by players. Bailey et al. (2020) demonstrated a model-based image-matching technique to calculate head kinematics. In their study, a ray-tracing technique was used to obtain the movement trajectories of the helmets. A helmet model was then manually superimposed onto the video images frame-by-frame to obtain the rotational movements of the helmets. Bailey et al found that cameras with a higher frame rate (240 Hz) are better for calculating the head kinematics than cameras with a lower frame rate (60 Hz), and their technique is more accurate at measuring the pre-impact velocities than during the impact. One of the shortcomings of this technique, when compared to the IMU-based sensor systems, is its reliance on a large amount of manual input, making it unsuitable for real-time monitoring in its current form. It is, however, still useful for capturing kinematics data for TBI studies and acquiring in-game helmet loading data for advising helmet testing.

While the aforementioned devices all come in different formats, they all aim to capture translational and rotational movements of the head, and impact locations. And as such, they all aim to assess the severity of head impacts under the same metrics. To rate the severity of a head impact, one needs to first be able to differentiate an actual impact from miscellaneous head accelerations in the kinematics data. To achieve that, many systems have simply set a threshold for the LA measurement (Siegmund et al. 2016, O'Connor et al. 2017, Wu et al. 2017). The threshold is often set to 10 g (ten times the gravitational acceleration), though other threshold values have also been suggested by researchers (Siegmund et al. 2016, O'Connor et al. 2017). This approach raises several concerns. When the threshold value is set too low, it

will result in a higher number of false positives, and when the value is set too high, some true positives will be missed out (Wu et al. 2017). In addition, since the cutoff is based on the LA, many impacts that are predominantly rotational with high AA values and low LA values will likely be missed out (Siegmund et al. 2016).

Two commercially available systems, HIT System and xPatch, have built-in software for identifying the legitimacy of a hit based on the acceleration waveform (Cumiskey et al. 2017). However, validation of this method is currently lacking in peer-reviewed literature. As an alternative, Wu et al. (2017) demonstrated a machine learning (ML) method for identifying head impacts in kinematic data. In their study, a support-vector-machine (SVM) classifier was trained using American football data collected on-field via IMU instrumented mouthguards. The trained SVM classifier achieved over 90% precision and recall when tested on an independent dataset. This shows promise as a more objective way to detect head impacts than arbitrary thresholding, and similar approaches can perhaps be implemented with future sensor systems to improve the accuracy of head impact detections.

Once an impact has been identified, the severity of the individual impact can be assessed. There has also been strong evidence in the literature that the accumulation of repeat head impact exposures can lead to certain risks (Nauman et al. 2015, Bari et al. 2019, Caccese et al. 2019, Mihalik et al. 2020), though there is still a large gap in that knowledge. As of now, the majority of head impact severity models only account for individual impacts. The most basic models are simply thresholds for either the peak LA or the peak AA at values that present a risk of concussion. These models are, however, extremely unreliable. At greater than 50% correct injury prediction levels, for every correct prediction made, hundreds if not thousands of false positives are also flagged (Greenwald et al. 2008). An NFL study has also attempted to correlate the peak LA to head injury rate using on-field data (Pellman & Waeckerle 2003). This paper claims that 75% of all impacts greater than 98.9 g would result in a concussion.

However, an independent study by Greenwald et al. (2008) found that out of the 3476 impacts of greater than 98.9 g they have recorded, only 11 of them (0.3%) have resulted in clinically diagnosed concussion. While the discrepancy could have been caused by data biased towards injurious impacts, it also indicates that it is likely that there isn't a clear TBI threshold.

Through a series of experiments carried out with human and dog cadavers at Wayne State University, it was found that head injury, defined as the occurrence of a skull fracture, correlated with the magnitude of LA and exposure duration. These data were plotted up as the famous wayne state tolerance curve (WSTC), see Figure 2.2 (Hutchinson et al. 1998). In 1966, Charles W Gadd analyzed this set of data, and by plotting the data on a log paper, fitted the data to the following equation,

$$\bar{a}^{2.5} \cdot T = 1,000 \quad (2.1)$$

where $\bar{a}$ is the average acceleration in g-force, and T is the exposure duration in seconds (Gadd 1966).

Gadd then expressed this fit in an integral form

$$I = \int_{\Delta t} a(t)^{2.5} dt \quad (2.2)$$

and called it the Gadd severity index (GSI). Later, out of consideration for the appropriateness of the GSI, Versace re-expressed Gadd's fit of the WSTC as

$$HIC = \left[\frac{\int_{t_1}^{t_2} a(t) dt}{(t_2 - t_1)} \right]^{2.5} (t_2 - t_1) \quad (2.3)$$

where HIC stands for Head Injury Criterion (Versace 1971). Although both GSI and HIC were developed for predicting life-threatening head injuries, they are, to

date, two of the most commonly used measures of impact severity in sports science for studying mild TBIs. While they have had great implications for regulations in the automobile industry, they have not proven to be any more reliable, statistically, than peak LA or RA thresholding for TBI prediction (Greenwald et al. 2008).

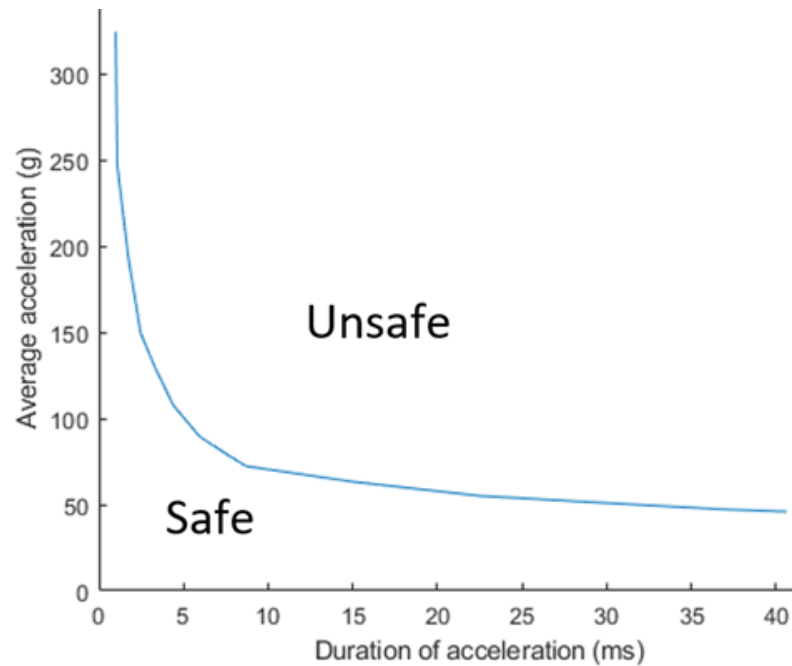


Figure 2.2: Wayne state tolerance curve (WSTC). Reproduced from Namjoshi et al. (2013)© 2013. Published by The Company of Biologists Ltd. CC BY 3.0

There are many criticisms of GSI and HIC, and many alternative measures of impact severity have been developed to address these criticisms. One of these criticisms is that neither GSI nor HIC accounts for rotational motion. As such, models that combine both linear and rotational measurements have been proposed. Such models include Kleiven's linear combination (KLC), a linear combination of HIC and the maximum resultant angular velocity (Kleiven 2007); the generalized acceleration model for brain injury threshold (GAMBIT) which regards LA and AA as proxies for stresses on the brain (Newman 1986); and the combined probability of concus-

sion (CP), a function of LA and AA developed from a head-impact dataset using multivariate logistic regression analysis (Rowson & Duma 2013).

Another criticism of GSI and HIC is that these models were loosely empirically fitted by hand and the dimensions of these measures are physically meaningless (Newman & Shewchenko 2000). Newman et al proposed to instead measure head impact severity using the rate of change of kinetic energy of the head and created the head impact power (HIP) index (Newman & Shewchenko 2000). While HIP is physically more sound than some of the other head impact measures, it still neglects some physical properties of the head. The human head is not isotropic and therefore responds differently to different directions of impact. A modified HIP was proposed to address this issue. The modified HIP has an additional scaling coefficient for each component of the original HIP, accounting for the various directions and differentiation between positive and negative accelerations (Kleiven 2003).

Greenwald et al. (2008) had also proposed a kinematics-based head impact measure that considers the anisotropic properties of the head, though via a different approach. By using principal component analysis, Greenwald et al constructed a head impact score that is a combination of GSI, HIC, LA, and AA. This score is then multiplied by a weighting factor based on the location of the impact to give what Greenwald et al called a weighted principal component score (wPCS). The wPCS is sometimes also referred to as the HIT severity profile (HITsp) since it was developed using the HIT system. Greenwald et al demonstrated that HIPsp performs better than the classical measures (LA, AA, and HIC) in predicting mild TBI. However, the prediction power of the HITsp is still insufficient for monitoring head traumas in practice as it has a positive predictive value (PPV) of 0.9% at the 50% sensitivity threshold (Greenwald et al. 2008). (The definitions of PPV and sensitivity are as follows. A 75% PPV at a threshold means that on average three of four impacts above this threshold result in mild TBIs. A 75% sensitivity at a threshold means that on average three of four

mild TBIs are resulted from impacts above this threshold.) Greenwald et al. (2008) also noted that a 'safe' HITsp value for one player may correspond to a mild TBI diagnosis for another player. These findings showcase the problem with regards to a 'one-size-fits-all head injury threshold'.

Studies have shown that rotational movements contribute more to brain injuries than linear motions (Ji et al. 2014). As a result, some head impact measures were developed while only considering rotational kinematics. Takhounts et al introduced the Brain Injury Criterion (BRIC) which is the sum of normalized maximum angular velocity and acceleration, and later introduced an updated version (BrIC) which only uses angular velocity after finding a low correlation between angular acceleration and strain on the brain (Takhounts et al. 2011, 2013). Later studies have, however, found the contrary and shown BRIC and BrIC to perform poorly under certain impact conditions (Gabler et al. 2016, Shi et al. 2020). Ommaya et al (Ommaya et al. 1967) constructed a tolerance curve for angular acceleration using data on concussed monkeys. The acceleration tolerance curve resembles the WSTC, and using this similarity, Kimpara & Iwamoto (2012) proposed two head impact measures similar to HIC. The first of the two is the power rotational head injury criterion (PRHIC) which replaces the LA term in HIC with the rotational component of HIP. The second is called rotational injury criterion (RIC) which replaces the LA term in HIC with AA. While both measures produced promising results on the initial study, the authors of RIC and PRHIC are concerned with the unusual physical dimensions of these measures and their implications on the physical interpretation. There are many more kinematics-based measures of head impact severity, though the sentiment here holds. Appendix A lists some of the commonly used kinematics-based measures of head impact severity within the field.

Despite all these efforts to establish new kinematics-based head impact severity measures and an injury threshold, no consensus has been reached within the field (Ji &

Zhao 2015). New sensor systems have been developed to capture a wider range of data in hopes of further improving upon current head impact severity measures. Merrell et al. (2017) demonstrated a helmet-based sensor system fitted with an array of nanocomposite foam (NCF) sensors. The NCF sensors can measure impacts through a triboelectric response. Merrell et al demonstrated that the NCF sensor system can accurately measure the impact location as well as the head kinematics. In addition to the information that conventional helmeted systems can capture, the NCF sensor system can also capture the impact energy. Future studies could investigate the relationship between impact energy and concussion, and potentially use the impact energy measures to improve head injury prediction accuracy.

2.1.6 Further considerations

With the growing public awareness of the short and long-term consequences of various sports-related injuries, the need to objectively safeguard and manage the athletes is clearer now, more than ever. There is a growing interest in understanding the well-being and mental health of athletes, which is affected by the ability to cope with both impacts and workload. A more holistic view is appropriate in the management of athletes and this indicates the necessity of a multi-modal approach for monitoring. At the moment there is a lack of research that aims to monitor across domains (e.g. impact, workload and well-being), which makes it hard to generate better models for player management. However, this approach also requires the individual tracking components to be well understood.

This review has identified a few issues with on-field monitoring that need to be addressed moving forward. The first issue is a general lack of data to help build and validate injury rate prediction models. There appear to be a lot of models in sports monitoring that were created from limited-sized and homogeneous data sets. With developments of better sensor systems, high-quality data needs to be captured

both in higher quantity and from a wider range of demographics to create models that are more accurate and more applicable across the sports community. Another issue is that there is a lack of consensus on how measurement data should be processed and presented. Inconsistent data presentations across related studies make comparing and combining outcomes of different studies difficult, and this may result in some researchers' hard work becoming much less impactful than they intended. Moving forward, standardized definitions and data formats that can facilitate meaningful comparisons between studies and meta-analyses are needed for the field to grow meaningfully and effectively. And lastly, there is a lack of diversity in physiological data that can be captured unobtrusively on-field. As things currently stand, two clinically-well-understood vital signs, temperature and breathing are rarely used in on-field monitoring. Temperature monitoring is currently done with expensive single-use ingestible thermometers. Respiratory measurements are mostly captured in lab settings in sports using bulky metabolic machines as a way to access the overall fitness of athletes rather than continuously monitored on-field.

2.1.7 Conclusion

This review provides an overview of on-field sports monitoring techniques that aim to improve the health and wellness of athletes. Most of the current monitoring methods are either tracking the athlete's workload or monitoring physical impact. Current sports monitoring practices are being held back by a lack of standard metrics and definitions, and the size and diversity of data needs to increase to develop ecologically valid approaches. The future of on-field sports monitoring will most likely consist of well packaged wearable sensor systems that can measure both physical and physiological parameters. One important physiological parameter that is currently under-utilized for on-field sport monitoring is breathing. Thus, it was decided that the research detailed in this thesis would aim to address this gap.

2.2 An Umbrella Review of Wearable Sensors For Respiratory Monitoring

2.2.1 Introduction

Building on the preceding scoping review of on-field monitoring techniques in sport, breathing was identified as a promising yet under-explored marker of internal workload. While current athlete monitoring practices are largely dominated by measures of external workload, impact exposure, heart rate, and subjective ratings of perceived exertion, respiratory activity has received comparatively little attention in on-field settings. This is notable because breathing reflects the integrated responses of multiple physiological systems to exercise and may provide valuable insight into exertion, fatigue, recovery, thermal strain, and overall physiological status (Nicolò et al. 2016, 2017, de Almeida e Bueno et al. 2021). As such, respiratory monitoring has the potential to complement existing workload metrics and contribute to a more comprehensive understanding of athlete health and performance.

Traditionally, respiratory assessment in sport has been conducted in controlled laboratory environments using metabolic carts and other bench-top systems. Although these methods can provide detailed information on ventilation and gas exchange, they are often impractical for continuous use during training and competition in a contact-sport context due to their cost, bulk, and restrictions on natural movement. Recent advances in wearable sensing technologies, however, have created new opportunities for ambulatory breathing monitoring. Various wearable sensor solutions have been developed for respiratory monitoring in medical applications such as sleep and respiratory health monitoring, as well as in sports performance tracking for individual sports such as running and cycling. Several review articles have examined different wearable approaches to breathing monitoring. This umbrella review was undertaken

to provide a comprehensive overview of the wearable technologies currently available for breathing monitoring.

2.2.2 Methods

The literature search was performed using the Web of Science database. The search strategy was designed to identify review articles relating to wearable technologies for respiratory monitoring. The following search query was applied: ("wearable*" [All Fields]) AND (("breath*" OR "respiratory" OR "respiration" OR "Tidal Volume" OR "Minute Ventilation" OR "VO2") [Title]) AND ((monitor*) OR (measure*) [Topic]).

The search was limited to review articles published before 2026. Articles retrieved from the search were screened for relevance using their titles and abstracts, with full-text assessment undertaken where required, to determine eligibility for inclusion in the umbrella review.

The final set of included review articles was analyzed to provide a high-level synthesis of the wearable technologies available for breathing monitoring. Data of interest included the types of wearable sensors discussed, the contexts in which they were applied, and the respiratory parameters they were designed to monitor.

2.2.3 Findings & Discussions

The database search identified 49 records from Web of Science, published between Jun 2017 to Dec 2025. Of these, 15 review articles met the inclusion criteria and were included in this umbrella review. Table 2.2 shows the list of review articles included in this umbrella review on wearable techniques for respiratory monitoring.

Table 2.2: Included review articles on wearable respiratory monitoring.

Title	Authors	Year
Wearable devices for respiratory monitoring	Vicente, Bernardo A Sebastiao, Raquel Sencadas, Vitor	2024
Artificial intelligence-based algorithms and healthcare applications of respiratory inductance plethysmography: A systematic review	Rahman, Md Shahidur Chowdhury, Sowrav Rasheduzzaman, Mirza Doulah, Abul Barkat Mollah Sayeed Ud	2024
Recent developments in wearable breath sensors for healthcare monitoring	Kim, Dohyung Lee, Jinwoo Park, Moo Kyun Ko, Seung Hwan	2024
Wearable respiratory sensors for health monitoring	Yin, Zhifu Yang, Yang Hu, Cong Li, Jinzhe Qin, Boyu Yang, Xue	2024

Continued on next page

Title	Authors	Year
Advances in respiratory monitoring: A comprehensive review of wearable and remote technologies	Vitazkova, Diana	2024
	Foltan, Erik	
	Kosnacova, Helena	
	Micjan, Michal	
	Donoval, Martin	
	Kuzma, Anton	
	Kopani, Martin	
	Vavrinsky, Erik	
Wearable sensors for respiration monitoring: A review	Hussain, Tauseef	2023
	Ullah, Sana	
	Fernández-García, Raúl	
	Gil, Ignacio	
Respiratory monitoring: Current state of the art and future roads	Costanzo, Ian	2020
	Sen, Devdip	
	Rhein, Lawrence	
	Guler, Ulkuhan	
Soft wearable sensors for monitoring symptoms of COVID-19 and other respiratory diseases: A review	Liu, Yuxuan	2022
	Shukla, Darpan	
	Newman, Holly	
	Zhu, Yong	
An overview of wearable piezo- resistive and inertial sensors for respiration rate monitoring	De Fazio, Roberto	2021
	Stabile, Marco	
	De Vittorio, Massimo	
	Velázquez, Ramiro	
	Visconti, Paolo	

Continued on next page

Title	Authors	Year
Sensing systems for respiration monitoring: A technical systematic review	Vanegas, Erik Igual, Raul Plaza, Inmaculada	2020
Recent development of respiratory rate measurement technologies	Liu, Haipeng Allen, John Zheng, Dingchang Chen, Fei	2019
Contact-based methods for measuring respiratory rate	Massaroni, Carlo Nicolò, Andrea Lo Presti, Daniela Sacchetti, Massimo Silvestri, Sergio Scheda, Emiliano	2019
Wearable technology: role in respiratory health and disease	Aliverti, Andrea	2017
State of the art review of wearable devices for respiratory monitoring	Jia, Zhiyu Huth, Hannah Teoh, Wen Qi Xu, Sheng Wood, Bradford Tse, Zion Tsz Ho	2025
Assessing the Tidal Volume through Wearables: A Scoping Review	Monaco, Vito Stefanini, Cesare	2021

Based on evidence synthesized from the 15 review articles found, wearable systems

for respiratory monitoring can be broadly classified into the following categories: techniques based on modulation of other physiological signals; techniques that measure chest wall volume changes and movement; techniques that measure properties of the air flow; and techniques based on acoustic sensing. Table 2.3 shows a breakdown of the wearable sensing techniques described in the included review articles coded into these categories.

Table 2.3: Coding of included review articles by wearable respiratory sensing technique category.

Article	Modulation of other physiological signals	Chest wall volume changes and movement	Air flow properties	Acoustic sensing
(Vicente et al. 2024)		✓		✓
(Rahman et al. 2024)		✓		
(Kim et al. 2024)			✓	
(Yin et al. 2024)			✓	
(Vitazkova et al. 2024)	✓	✓	✓	✓
(Hussain et al. 2023)		✓	✓	
(Costanzo et al. 2020)	✓	✓	✓	✓

Continued on next page

Article	Modulation of other physiological signals	Chest wall volume changes / movement	Air flow properties	Acoustic sensing
(Liu et al. 2022)	✓	✓	✓	
(De Fazio et al. 2021)		✓		
(Vanegas et al. 2020)	✓	✓	✓	✓
(Liu et al. 2019)	✓	✓	✓	✓
(Massaroni et al. 2019)	✓	✓	✓	✓
(Aliverti 2017)	✓	✓		
(Jia et al. 2025)		✓		
(Monaco & Stefanini 2021)		✓		✓

2.2.3.1 Modulation of other physiological signals

This category includes systems that do not measure breathing directly, but instead infer respiratory activity from changes imposed on other physiological signals. The most common examples are ECG and PPG, where respiratory rate or breathing pattern is estimated from cyclic respiratory modulation of cardiac electrical activity,

pulse amplitude, or pulse frequency. The same principles have also been applied to ballistocardiogram (BCG), seismocardiogram (SCG), osillometric cuff pressure pulses (OscP), and Korotkoff sounds (KorS) (Liu et al. 2019).

Indirect measurement of respiratory rate via other physiological signals has the advantage of providing an additional output from sensor systems primarily designed to capture another physiological signal, enabling multimodal sensing. However, these systems also have some distinct disadvantages when compared to the other measurement techniques. Extraction of accurate respiratory information can be difficult due to the complexity of the signals, and the accuracy of the respiratory signal can be highly sensitive to movement (Massaroni et al. 2019, Vitazkova et al. 2024). Due to these limitations, these techniques are better suited to sedentary monitoring applications, such as sleep monitoring or intensive care.

2.2.3.2 Chest wall volume changes and movement

This was the most dominant category across the included reviews. These systems estimate breathing by detecting thoracic and/or abdominal expansion, contraction, displacement, or deformation during the respiratory cycle. Common sensor systems include piezoresistive, fibre-optic, and capacitive strain sensors; respiratory inductance plethysmography; bioimpedance-based systems; and inertial sensors, integrated into belts, patches, textiles, or garments. (Aliverti 2017, De Fazio et al. 2021, Hussain et al. 2023, Rahman et al. 2024, Vicente et al. 2024, Jia et al. 2025, Massaroni et al. 2019).

These systems are primarily used to derive respiratory rate but have also been shown to estimate respiratory volume (De Fazio et al. 2021, Vitazkova et al. 2024, Monaco & Stefanini 2021). Due to the relatively unobtrusive nature of these systems, they are suitable for continuous monitoring of a subject as they go about their day, including during physical activities (Vanegas et al. 2020, De Fazio et al. 2021,

Vitazkova et al. 2024, Massaroni et al. 2019). However, chest-wall sensor systems may not be the most suitable form of wearable for contact-sport applications, where physical movements can introduce considerable artifacts into the signal (Lanata et al. 2009, Zhang et al. 2012, Vanegas et al. 2020), compromising sensor accuracy. Algorithms and sensor array solutions have been developed to address this issue, though it is noted that most solutions in the literature focus on walking and running. In the context of contact sports, it is uncertain how motion artifacts can affect the accuracy of these systems. And there is the additional concern that certain types of physical contact/collision can shift the sensors out of alignment. User comfort can also be a limiting factor for belt systems in contact sports, though smart garments are better suited for long periods of wear during exercise.

2.2.3.3 Properties of the air flow

This category includes systems that sense the physical or chemical characteristics of inspired or expired air. Reviews in this group describe airflow, temperature, humidity, and gas or chemical sensors positioned near the nose or mouth, either incorporated into masks or applied as stick-on patches. These systems may detect direct airflow via a turbine (Massaroni et al. 2019), or detect breathing through cyclic changes in air temperature or relative humidity, or by monitoring the composition of exhaled breath, including CO₂, O₂, and various other compounds expelled by the respiratory system (Kim et al. 2024, Yin et al. 2024). Currently, systems that monitor breathing-induced changes in temperature and humidity are best suited to controlled lab environments, given their form factors and susceptibility to environmental factors (Vitazkova et al. 2024, Liu et al. 2022). Turbine systems and gas sensors are less prone to motion and environmental artifacts and offer high measurement accuracy, but require bulky, intrusive equipment, including tightly strapped silicone face masks, to function (Massaroni et al. 2019), making them less ideal for real-time on-field monitoring.

2.2.3.4 Acoustic sensing

Respiratory parameters can be extracted from breathing sounds using advanced signal processing techniques. Acoustic sensing can be implemented using microphones or contact acoustic transducers placed along/near the respiratory tract or inside the ear canal. These systems are generally used for respiratory-rate estimation, breath-phase detection, and in some cases recognition of abnormal respiratory events such as wheeze, cough, or disordered breathing (Massaroni et al. 2019, Costanzo et al. 2020, Liu et al. 2019, Vitazkova et al. 2024). And with low noise data, acoustic sensors can also be used to measure tidal volume (Monaco & Stefanini 2021). The accuracy of acoustic respiratory sensing systems largely depends on the signal-to-noise ratio, which is influenced by sensor placement and noise isolation (Costanzo et al. 2020, Massaroni et al. 2019, Vitazkova et al. 2024). Acoustic sensors offer a relatively affordable and unobtrusive way to capture respiratory signal. Given their widespread integration into everyday devices such as headphones and smartphones, existing research has explored the use of wireless headphones for monitoring breathing during daily exercise (Ren et al. 2021, Kumar et al. 2021). Additionally, the compact size and low energy consumption of MEMS microphones allow them to be integrated into small, unobtrusive wearable devices such as earbuds (Martin & Voix 2017). This flexibility allows for the possibility of seamless sensor integration into existing protective equipment for contact sports, such as mouthguards or head protection gear.

2.2.4 Conclusion

This umbrella review synthesizes evidence from 15 review articles published between 2017 and 2025 to provide a comprehensive overview of wearable sensor technologies for respiratory monitoring. The literature indicates that wearable respiratory monitoring is a rapidly advancing field, with systems categorized into four primary methodological approaches: indirect estimation from modulation of other physiological signals,

measurement of chest wall volume changes and movement, sensing of airflow properties, and acoustic sensing. Collectively, these approaches demonstrate that breathing can be monitored using diverse wearable formats, such as belts, garments, patches, masks, and earplugs.

Chest wall movement-based systems, including piezoresistive, fibre-optic and capacitive strain sensors, respiratory inductance plethysmography, bioimpedance systems, and inertial sensors, are the most consistently represented across the included reviews. These systems are the most mature wearable solutions for continuous respiratory monitoring. These methods are attractive because they can be integrated into relatively unobtrusive wearables and capture both respiratory rate and volume. Indirect approaches based on ECG, PPG, and related physiological signals enable multimodal sensing, but their respiratory outputs are generally more susceptible to motion artifacts and signal-processing challenges. Airflow-property methods, such as temperature, humidity, flowmetry, and gas-based sensing, provide a more direct representation of breathing and may facilitate analysis of breath chemistry. However, their form factor and environmental sensitivity may limit their utility in free-living or contact sport settings. Acoustic approaches show promise as compact, low-burden solutions for detecting respiratory rate, volume, breath phase, and abnormal respiratory events, although their performance depends on signal quality, sensor placement, and noise control.

Overall, the findings indicate that wearable respiratory monitoring has significant potential to expand beyond traditional medical applications, such as sleep and respiratory disease management, into areas including exercise monitoring, recovery assessment, and athlete surveillance. Despite this potential, the current literature is primarily focused on healthcare, laboratory, or low-movement contexts, with limited direct evidence addressing the practical requirements of contact sports. For sport-specific applications, key considerations include robustness to impact and movement,

stability of sensor placement, comfort, unobtrusiveness, and compatibility with existing equipment. These factors may restrict the immediate applicability of some current systems, particularly airflow-based and indirect monitoring methods, while highlighting opportunities for the development of more discreet and durable sensing strategies.

This umbrella review identifies acoustic sensing as a particularly promising direction for the present project. Compared with other methods, acoustic approaches offer a combination of small sensor size, low burden, and integration flexibility, which may be especially advantageous in contact-sport settings. The potential to embed acoustic sensors into existing protective equipment, such as mouthguards or headgear, presents an opportunity to capture respiratory data without significantly affecting athlete movement or comfort. Future research should therefore investigate acoustic sensing solutions in greater detail, with specific attention to practical integration, signal quality, and utility for monitoring breathing in applied sporting contexts.

2.3 Breathing Monitoring Using Microphones in Sport and Exercise: A Rapid Review

2.3.1 Introduction

Accurate assessment of breathing during physical activity may provide valuable insight into exercise intensity, fatigue, and recovery (Nicolò et al. 2017, de Almeida e Bueno et al. 2021). Microphones have the potential to offer a compact, unobtrusive, wearable solution for monitoring breathing during training, competition, and recovery. Building on the previous umbrella review, this rapid review examines the current state of respiratory monitoring using microphones in sport and exercise settings.

2.3.2 Methods

For this review, three databases, Web of Science, IEEE Xplore, and PubMed were searched using keywords connected by Boolean operators. The following search query was applied: ("breath*" OR "respiration" OR "respiratory") AND ("microphone*" OR "acoustic sensor*") AND ("sport*" OR "exercise*"). The keywords were searched in Topic Field for Web of Science, All Metadata for IEEE, and Title/Abstract for PubMed. All articles published before 2026 were included in this search. The retrieved articles were screened following the PRISMA flowchart.

This review examined the use of microphones for respiratory monitoring in sport and exercise applications. to investigate the use of microphones for respiratory monitoring in sport and exercise applications. The Population, Concept, Context (PCC) framework was used to determine the inclusion criteria.

- Population - Studies involving athletes, recreationally active individuals, healthy participants, or clinical populations undertaking sport or exercise.
- Concept - Studies that use microphone-instrumented wearable systems to mon-

itor respiratory parameters or respiration-related internal workload measures.

- Context - Studies conducted in sport, exercise, training, competition, recovery, or other physically active settings, including laboratory and field-based environments.

Only peer-reviewed original research journal articles and full conference papers published in English were included.

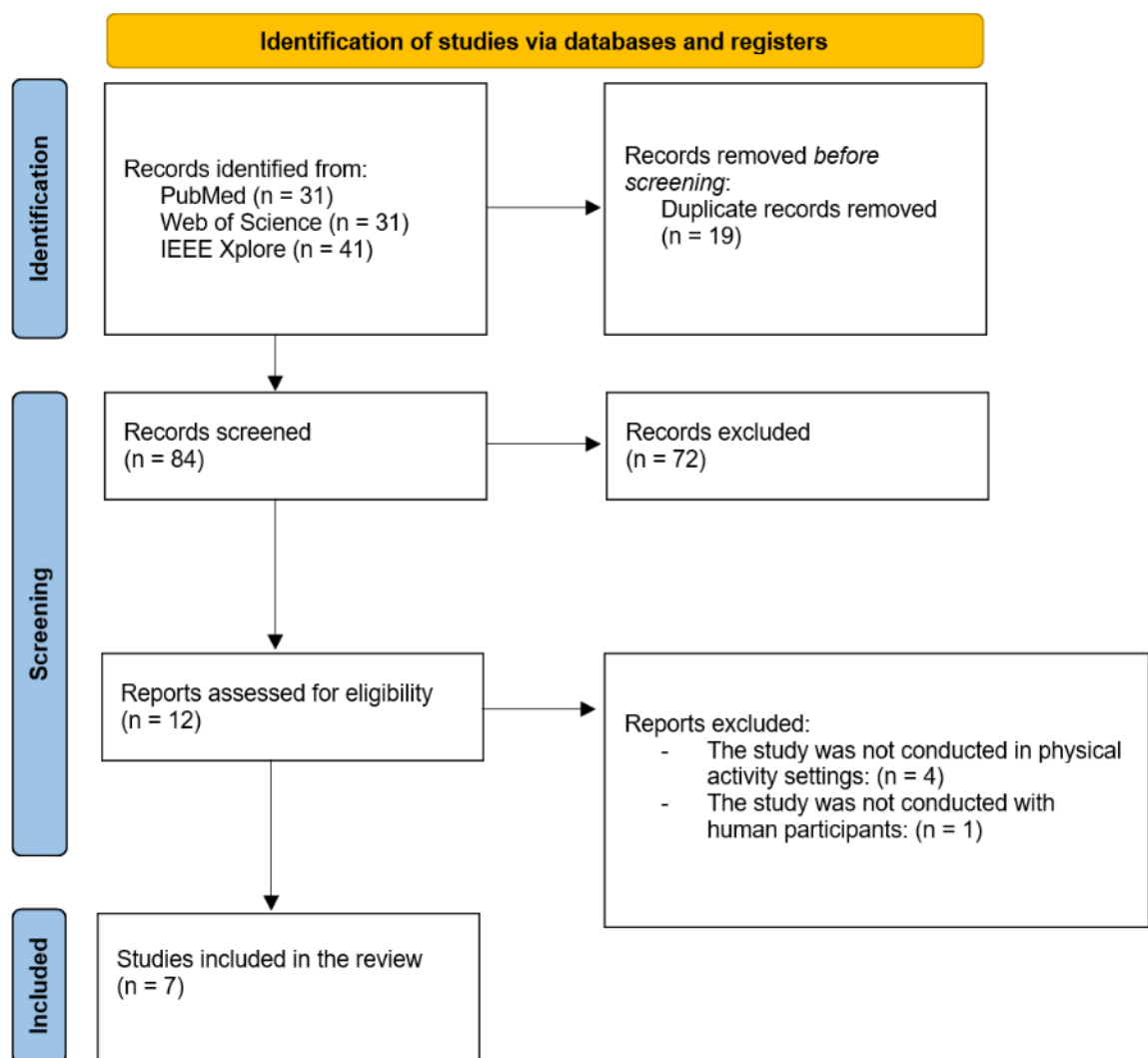


Figure 2.3: Article selection flow diagram, detailing the number of articles removed at each screening stage, and reasons for article removal.

2.3.3 Findings

The initial database search yielded 84 articles after removing the duplicates. Of these, 72 were excluded by screening the titles and abstracts. The full texts of the remaining 12 articles were examined, 5 of which were removed for failing to meet the inclusion criteria, leaving 7 eligible articles included in this review. Figure 2.3 shows the article selection flow diagram. Table 2.4 shows the list of included articles. Table 2.5 shows a breakdown of the included articles, with details on study design and applied technology.

Table 2.4: List of included articles

Title	Authors	Year
A Novel Audio-Perception-Based Algorithm for Physiological Monitoring	Z Zhang, W Jin, D Huang, Z Sun	2025
Breathing Sound-based Exercise Intensity Monitoring via Smartphones	Y Ren, Z Zheng, H Liu, Y Chen, H Li, C Wang	2021
BreathPro: Monitoring Breathing Mode during Running with Earables	C Hu, T Kandappu, Y Liu, C Mascolo, D Ma	2024
ESP32 Based Breathing Rate Monitoring System for Sport Activities	J Teličko, A Jakovičs	2024
Estimating Respiratory Rate From Breath Audio Obtained Through Wearable Microphones	A Kumar, V Mitra, C Oliver, A Ullal, M Biddulph, I Mance	2021
RespEar: Earable-Based Robust Respiratory Rate Monitoring	Y Liu, KJ Butkow, J Stuchbury-Wass, A Pullin, D Ma, C Mascolo	2025
Respiratory Rate Estimation during Walking and Running Using Breathing Sounds Recorded with a Microphone	C Romano, A Nicolò, L Innocenti, M Bravi, S Miccinilli, S Sterzi, M Sacchetti, E Schena, C Massaroni	2023

Table 2.5: A breakdown of study design and applied technology of included articles

Study	Wearable type	Sensor placement	Metric measured	Detection method	Activity level	Setting	n
Zhang et al. (2025)	Earable	In-ear	RR/ breathing phase	ML	Walking/Running	Outdoor	10
					Post workout recovery	Indoor	8
Ren et al. (2021)	Headphones	On the side of the head	Exercise intensity	ML	Running/Cycling/ Elliptical	Indoor (Lab)	9
Hu et al. (2024)	Earable	In-ear	RR/ breathing phase/ breathing mode	ML/Peak detection	Walking/ Running	Outdoor	25
Teličko & Jakovičs (2024)	Mounted headgear	In front of mouth and nose	RR	Peak detection	Walking/ Running	Indoor (Lab)	1
Kumar et al. (2021)	Headphones	On the side of the head	RR	ML	Walking/Running/ Biking/HIIT/ Calisthenics/Post workout recovery	Indoor/ Outdoor	21
Liu et al. (2025)	Earable	In-ear	RR/ breathing phase	Peak detection	Resting/Walking/ Running/Outdoor field work/Post workout recovery	Indoor/ Outdoor	18
Romano et al. (2023)	Facemask	In front of mouth and nose	RR/ breathing phase	Peak detection	Walking/Running/ Post workout recovery	Indoor (Lab)	10

The review found three common wearable form factors with three distinct sensor placements: wireless headphones with built-in microphones on the side of the head (Ren et al. 2021, Kumar et al. 2021); “earables” with microphones positioned within the ear-canal (Hu et al. 2024, Zhang et al. 2025, Liu et al. 2025); and instrumented facemasks/headgear with microphones positioned in front of the mouth and nose (Romano et al. 2023, Teličko & Jakovičs 2024). The most commonly measured metric is the respiratory rate (RR), with some studies going a step further to extract the

breathing phase in the time domain, enabling breath-by-breath detection. The most popular detection algorithm is peak detection, applied either to envelopes of wave amplitudes (Romano et al. 2023, Liu et al. 2025) or to time-frequency spectrograms (Hu et al. 2024, Teličko & Jakovičs 2024); machine learning algorithms are also present in the literature (Zhang et al. 2025, Kumar et al. 2021).

2.3.4 Discussions and conclusion

Although breathing monitoring via audio processing has been explored for decades (Massaroni et al. 2019), its application to sports monitoring is much more recent, as demonstrated by the earliest study included in this review, published in 2021. Earlier studies have used wireless headphones as a readily available wearable platform, leveraging their built-in microphones to capture breathing audio (Ren et al. 2021, Kumar et al. 2021). It was noted that this sensor placement did not isolate the microphone from environmental noise, and in high-background-noise conditions, breathing features in the recorded audio could become imperceptible, leading to measurement errors. This finding is consistent with those reported in Chapter 3 of this thesis. Due to this issue, Ren et al. (2021) opted to directly estimate exercise intensity from breathing audio using a Gaussian mixture model, circumventing RR detection.

A natural progression from headphone-based respiratory monitoring has been the move towards in-ear sensing. Recent advances in headphone technology, particularly the widespread availability of in-ear active noise-cancelling devices with both external and in-ear microphones, have created new opportunities for respiratory audio monitoring with off-the-shelf wearables. Nicknamed “earables”, in-ear cardiorespiratory audio monitoring has previously been explored in non-sport settings as a feasible approach for continuous monitoring of breathing rate and heart rate (Pressler et al. 2004, Kusche et al. 2015, Martin & Voix 2017). In the context of exercise, in-ear devices offer practical advantages over conventional headphone-based configurations,

particularly for environmental noise isolation (Liu et al. 2025). In addition to better noise isolation, Hu et al. (2024) demonstrated that the external microphone can be used to capture background noise, which can then be subtracted from the in-ear signal to further reduce noise in the cardiorespiratory audio.

While earables have addressed the background-noise issue faced by external microphones, they also present a unique challenge: heartbeat sounds can be entangled with respiratory sounds, especially at low physical activity levels when breathing is quiet and shallow. Two studies, Zhang et al. (2025) and Liu et al. (2025), proposed algorithms to decouple heartbeat signals from breathing signals, enabling measurement of both heart rate and RR. Though Liu et al. (2025) noted that heartbeat signals are only distinguishable when the user is not physically active. On the other hand, Hu et al. (2024) recommended a high-pass filter at 500Hz, which minimizes body-generated sounds such as heartbeat sounds.

Another wearable approach to breathing monitoring using a microphone reported in the literature is to place the microphone directly in front of the mouth and nose, either mounted on a headgear (Teličko & Jakovičs 2024) or attached to a facemask (Romano et al. 2023). Both studies found that the approach performed with high accuracy in lab settings, but noted that error increased with ambient noise level. Additionally, Teličko & Jakovičs (2024) reported that the monitoring accuracy increased at higher activity levels, due to an increase in breathing intensity, a finding that's consistent with those reported in Chapter 5 of this thesis.

All the included studies acknowledged footsteps during walking and running as a source of noise in the breathing audio. This is universally addressed via high-pass filtering, though the threshold frequency varies across studies, ranging from 200Hz to 500Hz for breathing rate detection. Liu et al. (2025) also suggested a bandpass filter with cutoffs of 2000Hz to 9000Hz to leverage the harmonics of breathing sounds

during high-intensity running, where step sounds severely overwhelm the breathing sounds in their signal.

Peak-detection-based algorithms were found to be the most popular method for estimating breathing rate. With high-definition breathing audio, peak detection performed on either sound amplitude envelopes (Romano et al. 2023, Liu et al. 2025) or time-frequency spectrograms (Hu et al. 2024, Teličko & Jakovičs 2024) has been shown to be able to estimate breathing phase transition points. Hu et al. (2024) also presented an additional processing step after breathing phase estimation, where a KNN-based classifier is used to determine the mode of breathing (i.e. nose/mouth inhale/exhale). Kumar et al. (2021) and Zhang et al. (2025) each presented a deep learning model comprising CNN and RNN layers to extract the breathing rate from spectrograms of breathing audio, reporting accuracies similar to those of the peak detection approach.

Overall, the evidence in the literature suggests that microphones are a feasible sensing option for respiratory monitoring in sports. Though it is noted that the existing wearable solutions are ideal for non-contact individual sports such as running or cycling, they are not suited to team contact sports such as hockey or rugby. There's a need to develop a wearable form factor tailored to the needs and constraints of team contact sports. It is also noted that the current literature is heavily focused on respiration rate detection. Future work should attempt to extract breathing volume from breathing audio as an additional metric.

2.4 Chapter Conclusion

The literature reviewed in this chapter establishes both the rationale for this thesis and the specific methodological direction it takes. First, the scoping review of on-field sports monitoring highlighted that current approaches to athlete health and wellness monitoring in team sports are dominated by measures of workload and physical impact. Although these methods provide valuable information, progress in the field is constrained by inconsistent metrics and definitions, as well as by a lack of sufficiently large and diverse datasets to support ecologically valid monitoring approaches. The review also indicated that the future of on-field athlete monitoring is likely to depend on integrated wearable systems capable of capturing both physical and physiological parameters. Within this context, breathing emerged as an important but under-utilized physiological signal, identifying a clear gap that this thesis seeks to address.

The subsequent umbrella review showed that wearable respiratory monitoring is a rapidly developing area, with four broad methodological approaches represented in the literature: indirect estimation from other physiological signals, chest wall movement sensing, airflow-property sensing, and acoustic sensing. Among these, chest wall movement-based systems appear to be the most mature and consistently represented, offering practical routes for continuous monitoring of respiratory rate and, in some cases, breathing volume. However, the broader evidence base remains centred on healthcare, laboratory, and low-movement applications, with limited attention given to the practical demands of contact-sport environments. For sporting use, especially in team contact sports, wearable respiratory systems must be robust to impact and movement, stable in placement, comfortable, unobtrusive, and compatible with existing equipment. These requirements reduce the suitability of many current solutions and point towards the need for sensing approaches that are smaller, more discreet,

and more easily integrated into sport-specific wearables.

From this broader landscape, acoustic sensing was identified as a particularly promising direction. The rapid review of microphone-based breathing monitoring demonstrated that audio approaches are feasible for exercise settings and can be implemented in several wearable formats, including headphones, earables, and microphones positioned near the mouth and nose. The literature shows that these systems can estimate respiratory rate with good accuracy under controlled conditions, typically using peak-detection or deep learning approaches applied to breathing audio. While demonstrating good potential for respiratory monitoring in non-contact sports such as running and cycling, current microphone-instrumented wearable form factors are incompatible with the demands of contact sports. There's a need to develop a wearable form factor tailored to the needs and constraints of team contact sports.

Taken together, the findings of this chapter show that breathing is a valuable and underexplored target for on-field sports monitoring, and that acoustic sensing offers a strong foundation for addressing this gap. However, they also make it clear that there is a need for a wearable respiratory monitoring approach specifically designed for team contact sports. In addition, the current literature remains heavily focused on respiratory rate, with comparatively little attention given to extracting breathing volume and other potentially meaningful metrics. Therefore, this thesis is positioned to contribute by investigating practical acoustic approaches to breathing monitoring in applied contact sport contexts, with particular emphasis on wearable integration, signal robustness, and the development of algorithms to extract respiratory measures (e.g. RR and VE) and related internal workload metrics relevant to athlete monitoring in team contact sports.

Chapter 3

Development And Preliminary Testing of An Oral Cavity Based Respiratory Monitoring Sensor System

Chapter Abstract

The previous chapter found that current on-field monitoring practices in team sports only capture a narrow range of physiological metrics. Respiratory metrics such as breathing rate and minute ventilation have been identified as promising metrics for field-based sports monitoring. While respiratory metrics are commonly used in lab-based metabolic tests, they have yet to be implemented on-field. Recognizing the potential value respiratory monitoring could offer in sports, this chapter describes the design, implementation, and preliminary validation of an acoustic sensing based respiratory monitoring system designed for use in contact-sport environments.

It was proposed that breathing can be captured on-field via acoustic sensors embedded inside existing sports gears. Two suitable locations for such a system have been identified, either as a smart mouthguard or an instrumented headgear. Four participants were recruited for the study due to COVID constraints, and respiratory data were captured in both indoor and outdoor settings.

The signal-to-noise ratio (SNR) at these placements and their potential for respiratory rate estimation were compared. A fast-Fourier transform (FFT) based frequency domain analysis was used to estimate the respiratory rate and compare the measurement accuracy of the two sensor configurations. It was found that the SNR of the oral placement is significantly better than the head placement for both indoor and outdoor settings. It was also found, through agreement analysis, that the oral placement had a significantly smaller error compared to the head location when predicting respiratory rate. Additionally, to evaluate the acoustic sensor's potential to capture respiration volume in addition to respiration rate, a correlation test was carried out between the breathing audio volume and minute ventilation measured via a metabolic cart. Results from the pilot study indicate that the mouthguard sensor placement can reliably capture respiratory activity under controlled conditions, supporting its feasibility as an unobtrusive continuous monitoring modality.

Furthermore, the chapter details the design considerations and system development. Practical constraints related to usability, safety, and integration into sporting contexts were addressed. The finalized mouthguard design and fabrication process are documented in this chapter.

3.1 Introduction

The literature review has found that the health and well-being of athletes can be improved via monitoring their internal loads, physiological responses to physical workloads. Within internal-load monitoring, only two methods are widespread, survey-based ratings of perceived exertion and heart-rate based metrics. And while RPE demonstrates a good representation of an athlete’s physical exertion, it is difficult to achieve continuous RPE monitoring due to its survey-based nature.

Emerging evidence suggests that respiratory parameters exhibit a stronger association with RPE than heart rate and may provide a more sensitive indicator of physiological strain and recovery status (Nicolò et al. 2016). Concurrently, prior research has demonstrated the feasibility of using microphones to capture respiratory signals through audio-based analysis (Nam et al. 2015, Martin & Voix 2017).

Accordingly, this pilot study was conducted to evaluate the feasibility of monitoring respiratory activity using microphones embedded within sports equipment. In addition to assessing signal viability, preliminary safety evaluations and device development processes were undertaken to ensure the practicality and usability of the proposed system.

3.2 Sensor Placement & Breathing Rate Detection

Studies have shown that microphones can be used to obtain respiratory information by applying audio analysis techniques (Nam et al. 2015, Martin & Voix 2017). One study has attempted to implement this idea in an industrial environment by embedding microphones into protective earbuds (Martin & Voix 2017). A similar approach can be taken in contact sports. Two suitable locations have been identified for microphone placement in contact-sports settings; (i) mounted to a piece of headgear such as a

pair of sport safety glasses, a helmet or a scrum cap (a sensor location that has been presented in the literature (Kumar et al. 2021)), or (ii) embedded inside a mouthguard. This pilot study aimed to find the more optimal microphone placements for measuring respiratory rates in contact-sport athletes between the two microphone configurations. Specifically, the signal-to-noise ratio (SNR) of the recorded audio data was evaluated, and a simple data-processing pipeline for respiratory rate estimation was developed and assessed across both microphone placements.

3.2.1 Methods

This study was conducted under the ethics approval R70833/RE001 granted by the Oxford University Medical Sciences Interdivisional Research Ethics Committee.

3.2.1.1 Data Acquisition

Four healthy participants, one female and three males, between the ages of 21 to 30 were recruited for this study. Two data capture sessions were performed with each participant individually, one session under a quiet indoor condition, and the other under an outdoor condition simulating a more noisy sports pitch. During each session, the volunteer was fitted with a customized mouthguard integrated with electronics. The mouthguards were thermoformed using the participants' dental impressions. The mouthguards were manufactured with an electronics enclosure extending behind the teeth and conformed to the palate surface, and inside the enclosure sits a STEVAL-STLKT01V1 development board (STMicroelectronics, Switzerland; see figure 3.1). The electronics inside the mouthguards are each equipped with a digital microphone (MP34DT05-A, STMicroelectronics, Switzerland) with a 64 dB signal-to-noise ratio and $-26 \text{ dBFS} \pm 3 \text{ dB}$ sensitivity, sampling at 8 kHz and 16 bits per sample. A second set of identical electronics were also attached to goggles, which were given to the participants to represent the headgear placement. And a third off-person set

of identical microphones were placed within 30cm of the participants to function as an external control set. This decision was based on the results reported by (Nam et al. 2015), which claimed that breathing sounds can be picked by a consumer-grade microphone placed as far as 30cm from a person. Figure 3.2 shows an example of this setup.



Figure 3.1: A STEVAL-STLKT01V1 development board

Each data capture session aimed to collect at least 10 minutes of breathing audio with occasional coughing and throat-clearing noises. During data capture, the participants were instructed to breathe naturally. The beginning of each clip was marked by the participant performing an audio synchronization action consisting of holding their breath, which helped to check synchronisation of the data captured by different microphones; they were also instructed to hold their breath at multiple instances throughout each data capture session, to collect normative background noise for the SNR analysis. The participant were also asked to log their breathing, either using a time-stamped key-logger program written in python on a laptop, or a phone app written in C# that does the same (see appendix B). These logs were then used as the ground truth during data analysis.



(a) Experimental Setup



(b) Custom Instrumented Mouthguard

Figure 3.2: An example of a data capture session. The participant is hooked up to three microphones, one strapped to their glasses, one in their mouth inside a mouthguard, and one on the park bench. The participant is seen logging their breathing on the computer using a python program.

3.2.1.2 Signal to Noise Ratio

The raw audio files acquired from the data capture were clipped into multiple clips of continuous breathing, each around one minute in duration. Eighty clips of one-minute-long continuous breathing audio were cropped from the raw data per microphone placement, 40 under indoor conditions and 40 for the outdoor condition. Open-source audio editing software, Audacity® (<https://github.com/audacity/audacity>) was used during the manual labelling of the data, which helped to label the data accurately down to milliseconds.

SNRs were calculated using the root mean squared values of the pure breathing clips and the average noise level obtained during the requested non-breathing episodes for each participant. The SNR is given by the following formula:

$$SNR = 20 \cdot \log\left(\frac{RMS_{signal}}{RMS_{noise}}\right) \quad (3.1)$$

3.2.1.3 Breathing Rate Detection

For breathing rate detection, the clips of continuous breathing were pre-processed through a series of filters. A 4th order Butterworth band-pass filter of 100Hz-400Hz was applied forward and back on the data to get rid of low and high-frequency noises outside of the range of breathing (de Almeida e Bueno et al. 2021). Following that, the data were downsampled to 1600Hz. Audio envelopes were then extracted via a Hilbert transformation followed by a moving average which subsequently further reduced the data to 8Hz.

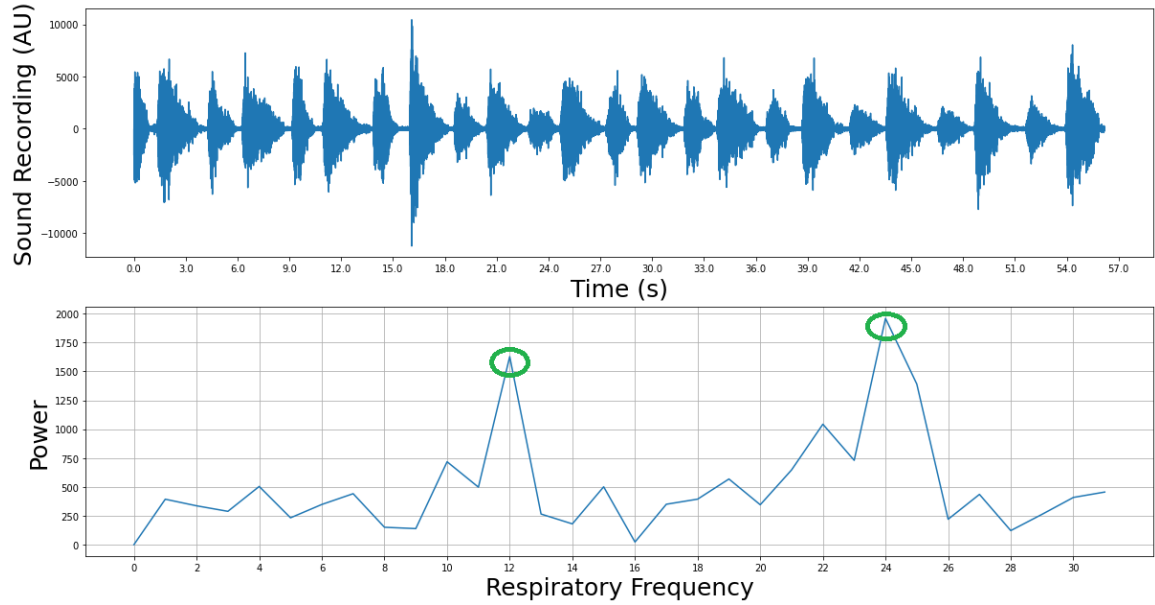


Figure 3.3: Top: an audio clip of breathing which contains 12 breaths; bottom: a Fast Fourier Transform of the audio clip. The peak with the highest power was found at 24 BPM, and a corresponding prominent peak was found at 12 BPM. The two peaks are offset by a factor of 2, indicating that the respiratory rate is 12 BPM, with at peak at 24 BPM representing the number of inhales and exhales combined.

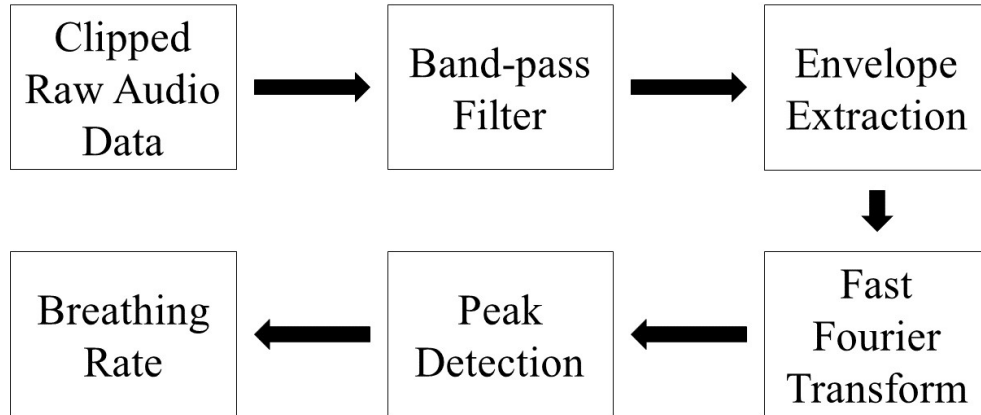


Figure 3.4: A simple block diagram of the respiratory rate estimation algorithm.

A FFT was performed on the sound envelopes to transform the data into the frequency domain. A hard boundary of 6 breaths per minute (BPM) to 31 BPM was set on the frequency domain analysis, based on clinical survey results on the breathing rate of adults at rest (Benchetrit 2000). A peak detection algorithm was then implemented to look for a pair of high-power peaks in the FFT separated by a factor of two, with the lower one representing the number of full breath cycles, and the higher one representing the number of inhales and exhales combined (Figure 3.3). The peak with the lower frequency value was then selected as the respiratory rate of the audio clip. Figure 3.4 shows a block diagram of the algorithm pipeline. The respiratory rate estimations were compared to the respiratory rate logged by the participants during the data collection for validation.

3.2.1.4 Analysis

Due to the small sample size of the pilot study, the data captured via the microphone configurations were compared descriptively. The mean SNR values for the microphone configuration under both indoor and outdoor settings were computed and directly compared side-by-side. Additionally, the quality of the wave forms of the audio sig-

nals for the different microphone configurations was compared qualitatively. For the breathing rate detection accuracy, the mean percentage error as well as the number of correctly predicted audio clips were compared across the different microphone locations. And lastly, the agreement between the predicted breathing rates and the recorded ground truths were compared between the mouthguard configuration and the headgear configuration under indoor and outdoor conditions using Bland-Altman plots.

3.2.2 Results

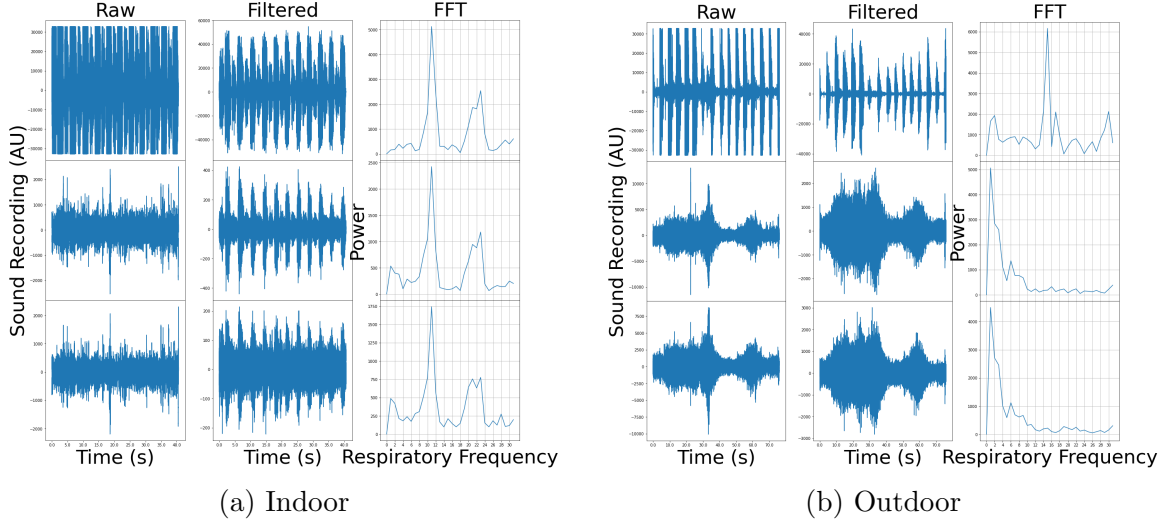


Figure 3.5: These graphs show samples of time series data captured by microphones placed in different locations to visually demonstrate the difference in data quality. The top row shows mouthguard data, the middle row shows headgear data, and the bottom row shows external control data. It is visually clear that mouthguard data contain the least amount of noise and have the best signal strength.

The breathing clips included a mix of mouth and nose breathing. The respiratory rates contained within these clips ranged from 7 to 19 breaths per minute (BPM), with an average of 12.6 BPM (root mean squared deviation, $\text{RMSD} = 3.3$ BPM),

all of which fall well within the 6 BPM to 31 BPM range found in the literature. The mouthguard-based microphone out performed the headgear-based microphone as well as the control microphone under both indoor and outdoor conditions, while the headgear-based microphone and the control performed similarly. Figure 3.5 demonstrates the difference in data quality between different microphones under both indoor and out door conditions, showing that the mouthguard configuration can capture higher quality breathing data. The data quality difference is especially clear under noisy outdoor conditions.

The mouthguard-based microphone has an average SNR of 16.04dB across all four subjects under indoor conditions, while the headgear-based microphone and the control have an average SNR of 4.42dB and 3.63dB respectively. Under outdoor conditions, the mouthguard-based microphone has an average SNR of 14.96dB, the headgear-based microphone has an SNR of 2.09dB, and the control has an average SNR of 1.93dB.

The breathing rate extraction algorithm performed the best with the mouthguard-based microphone dataset. Over all 80 minute-long clips, a mean percentage error of 4.02% was found, with 51 (63.75%) correct predictions and 75 (93.75%) predictions within 1 BPM. Table 3.1 shows a detailed breakdown of the results from the breathing rate algorithm, the errors found in each dataset under both indoor and outdoor conditions. Figure 3.6 shows Bland-Altman plots displaying the agreement between the predicted breathing rate and the manually logged breathing rate for the mouthguard microphone configuration and the head gear microphone configuration, under indoor and outdoor conditions.

		Mouthguard-based Microphone	Headgear-based Microphone	Control Microphone
Overall n=80	Mean percentage error	4.02%	43.38%	41.89%
	Correct predictions	51 (63.75%)	10 (12.5%)	8 (10.0%)
	Predictions within 1 BPM	75 (93.75%)	18 (22.5%)	12 (15.0%)
Indoor n=40	Mean percentage error	3.47%	38.19%	37.08%
	Correct predictions	27 (67.5%)	8 (20%)	6 (15.0%)
	Predictions within 1 BPM	38 (95.0%)	13 (32.5%)	9 (22.5%)
Outdoor n=40	Mean percentage error	4.58%	48.56%	46.71%
	Correct predictions	24 (60.0%)	2 (5.0%)	2 (5.0%)
	Predictions within 1 BPM	37 (92.5%)	5 (12.5%)	3 (7.5%)

Table 3.1: Results from the breathing rate algorithm

3.2.3 Discussion

This study was conducted toward the end of the COVID-19 lockdown, which hindered the recruitment process. All four participants were members of the research group who already had a dental cast made. While the sample size is fairly small, it was deemed appropriate for pilot testing before committing to further research and development.

The oral-cavity sensor placement captures higher quality data than external sensor placements. A simple frequency-domain-analysis algorithm without any optimization was applied for BPM estimation to minimize biases towards a particular microphone placement. It was noted, qualitatively, that the estimations with higher errors resulted from audio data that were less uniform in either amplitude or breath duration. When comparing the Bland-Altman plots for the mouthguard configuration under indoor and outdoor conditions, there is no noticeable difference in the accuracy. This observation hints that the mouthguard configuration has the potential to collect high

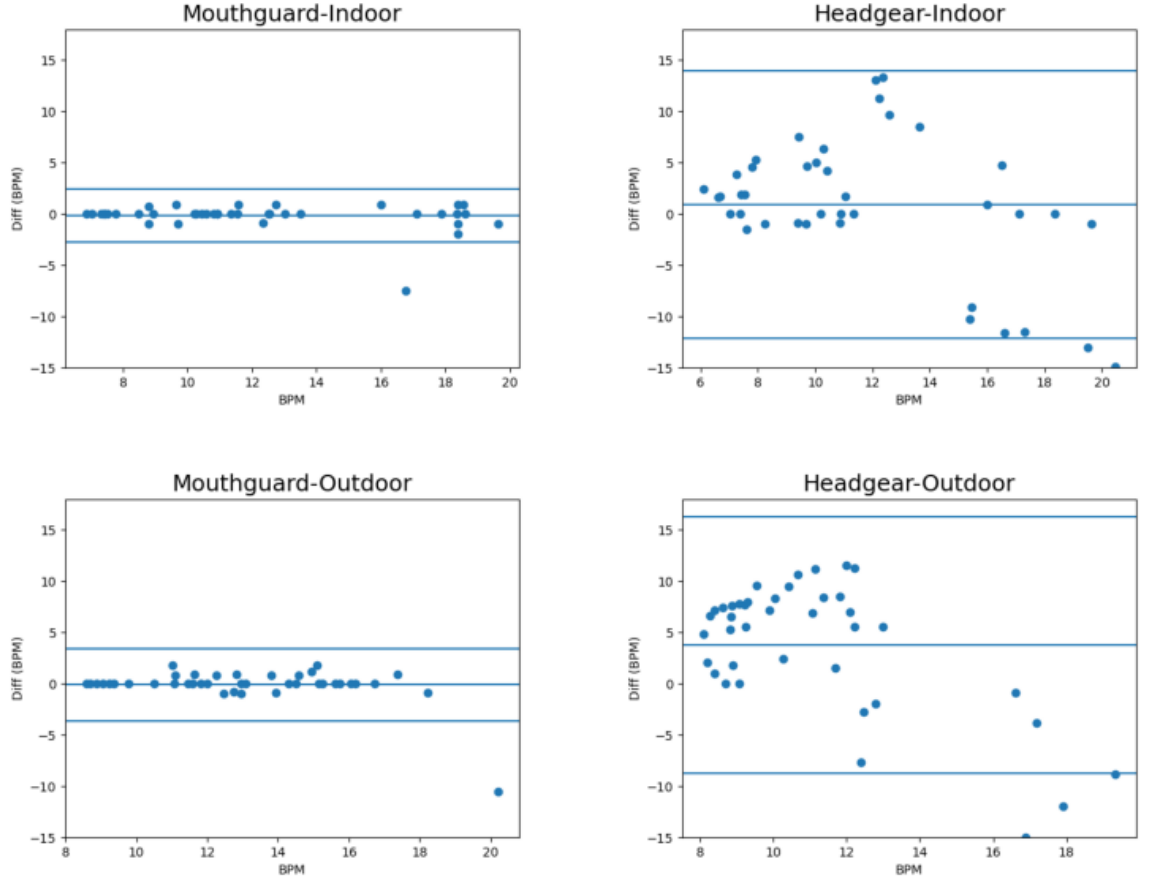


Figure 3.6: Bland-Altman plots displaying the agreement between the predicted breathing rate and the manually logged breathing rate for the mouth-guard microphone configuration and the head gear microphone configuration, under indoor and outdoor conditions. It is clear that the mouthguard configuration out performs the head gear configuration.

quality data even in a noisy environment.

The simple algorithm proposed in this study is similar to those found in literature (Martin & Voix 2017, Nam et al. 2015). Nam et al. (2015) demonstrated a similar breathing rate algorithm intended for a microphone placed within 30cm to a person. Nam et al’s algorithm follows a similar pre-processing pipeline as the one proposed

in this study, in which the data is band-passed and then passed through a Hilbert Transform based sound envelope extraction. A spectrogram of the envelope is then generated, and a peak detection is carried out on the spectrogram. Nam et al reported that their algorithm performs with good accuracy for breathing rates above 6 BPM and below 40 BPM, with a median error at around 30%. Martin & Voix (2017) presented a similar but more direct approach where a peak detection is applied directly to the sound envelopes to extract the breathing rate under 30 BPM. Martin and Voix implemented their algorithm with a ear-canal based microphone intended for industrial settings. They reported a mean percentage error of 30.9%, and commented that the algorithm under-performs in loud environments. The results from these two studies are comparable to with the errors found in the control microphone data (41.89%) and the headgear-based microphone data (43.38%) of this study respectively. These results further prove that the oral cavity (with a mean percentage error of 4.02%) is the most suitable microphone placement location for extracting breathing rate, especially in a sports setting where the microphone can easily be integrated into a piece of existing protective gear, the mouthguard.

3.3 Using Audio Volume for Minute Ventilation and RPE Predictions

The acoustic wave energy of breathing audio can be used to model tidal volume during sleep and static tasks (Yadollahi et al. 2013). This suggests that minute ventilation (VE) - a product of tidal volume and breathing rate, can be estimated through a function of breathing rate and breathing audio volume. VE has been shown to correlate well to an athlete's rating of perceived exertion (de Almeida e Bueno et al. 2021). Thus an audio signal-based VE estimator was preliminarily explored.

3.3.1 Methods

This investigation uses previously published data from de Almeida e Bueno et al. (2021), which was conducted under the ethics approval R43470/RE001 granted by the Oxford University Medical Sciences Interdivisional Research Ethics Committee.

3.3.1.1 Data Acquisition

Eight healthy male participants were recruited for this study. The participants were all university level athletes from contact sports. The participants were between 20–39 years of age, weighed between 72–120kg, and between 170–187cm tall. Data collection sessions were performed outdoors, where the participants were asked to wear both an instrumented oral-cavity device for respiratory audio recording and a portable metabolic cart to record ground truth for breathing rate and minute ventilation. The participants were then asked to perform a 20m shuttle-run fitness test known as the beep test.

3.3.1.2 Data Processing

The audio data were filtered with an 4th-order zero-phase Butterworth 100Hz to 400Hz band-pass filter to remove background noise (de Almeida e Bueno et al. 2021). For the purpose of this analysis, an audio-based respiration estimator (aVE) is defined as the product of the moving average (with a 10s-wide sliding window) of the absolute breathing audio volume and the breathing rate obtained from the metabolic cart. This aVE value is tested for corelation against VE measured by the metabolic cart.

3.3.2 Results

The aVE was found to have a good Pearson’s r-value ($r = 0.88$) with actual VE collected by the metabolic cart. Figure 3.7 shows a comparison between VE and aVE. This indicates the potential to measure VE using breathing audio signal.

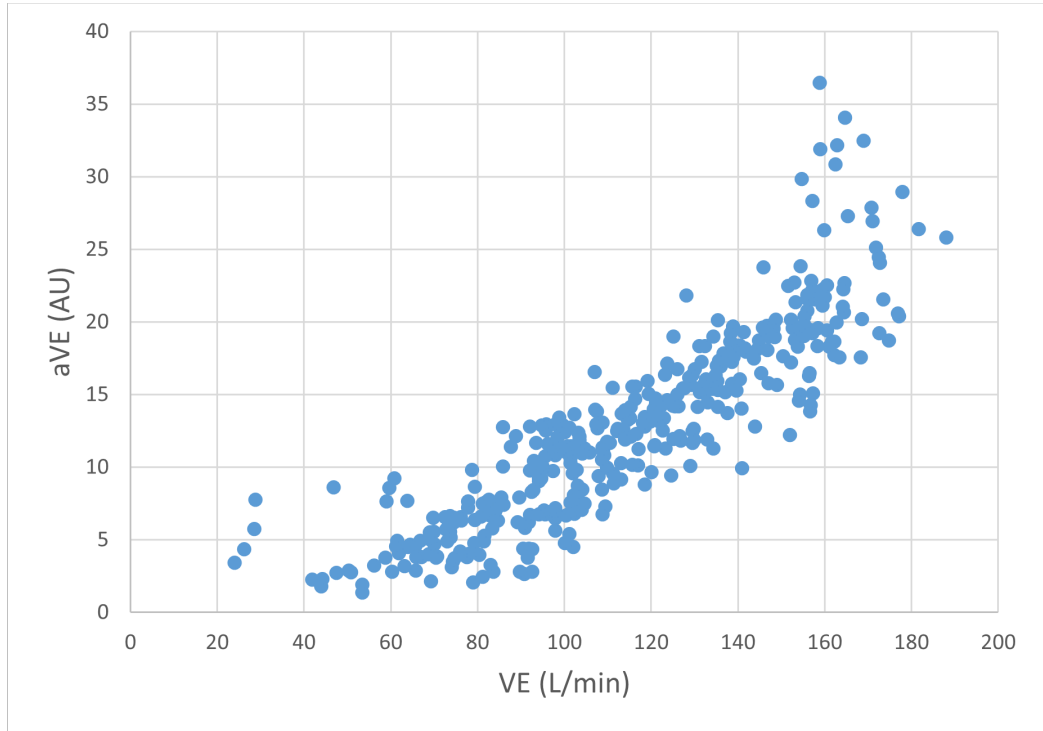


Figure 3.7: aVE, which is defined as the product of breathing rate and breathing audio volume, shows a strong correlation to VE, which is the volume of air exchanged during breathing per minute

3.4 Instrumented Mouthguard Development

After determining that a microphone-integrated mouthguard has the potential to capture respiratory data in sporting environments, a series of design and development work was undertaken within the Natural Interaction Lab to create an instrumented mouthguard that is safe and unobtrusive. The MP34DT05-A MEMS digital microphone on the development used in the initial pilot study demonstrated reliable performance when encapsulated in EVA, and was therefore retained in subsequent design iterations.

Although an in-house approach to custom electronics development was initially explored, this was ultimately outsourced due to time constraints. Consequently, the present work focused primarily on the safety and ergonomic optimisation of the de-

vice. A formal risk assessment identified the battery as the component posing the greatest risk, with three primary high-risk failure modes: mechanical damage resulting from impact or biting, water ingress leading to electrical short-circuiting, and exposure to elevated temperatures during manufacturing potentially resulting in battery malfunction. Design decisions were made to mitigate these risks.

3.4.1 Manufacturing Method Selection

Two custom mouthguard fabrication methods were identified through consultation with dental experts and a brief review of literature: the conventional approach, involving thermoforming EVA over a dental cast (Gould et al. 2019); and a more recent, experimental approach based on additive manufacturing of a mouthguard model derived from a 3D dental scan (Sousa et al. 2021). Water ingress was identified as one of the major risks in the risk assessment. Selection of an appropriate manufacturing method was considered critical in mitigating this risk. Thus, water ingress testing was undertaken to evaluate these two mouthguard manufacturing methods.

To assess the ability of each method to protect embedded electronics from moisture exposure, 3D-printed mock-ups representing the electronic components were produced and fitted with moisture indicator paper. For the thermoforming approach, five mouthguards—similar to the design shown in Figure 3.2b—were fabricated by thermoforming a 1.5mm EVA sheet over a dental cast, positioning the electronics mock-ups within the formed structure, and subsequently encapsulating them with an additional 3mm EVA layer thermoformed over the assembly. The 3mm thickness of the outer layer was determined by a previous study from the group, which found that 3mm thick EVA is sufficient to protect embedded electronics from forces experienced in contact sports (Bridgman et al. 2019).

To evaluate the performance of additively manufactured mouthguards, five each

of cuboid electronics enclosure units with 1mm, 2mm, and 3mm wall thicknesses were fabricated using biomedical-grade TPU (DSM N.V., Maastricht, Netherlands) via Fused Deposition Modeling (FDM). To embed the electronics mock-ups inside the enclosure, the printing process was intentionally paused at an intermediate stage to allow for the insertion of the electronics mock-ups, after which the print was resumed to fully encapsulate the components within (see figure 3.8). TPU was selected as the test material due to its comparable impact absorption properties to EVA (Sousa et al. 2021).

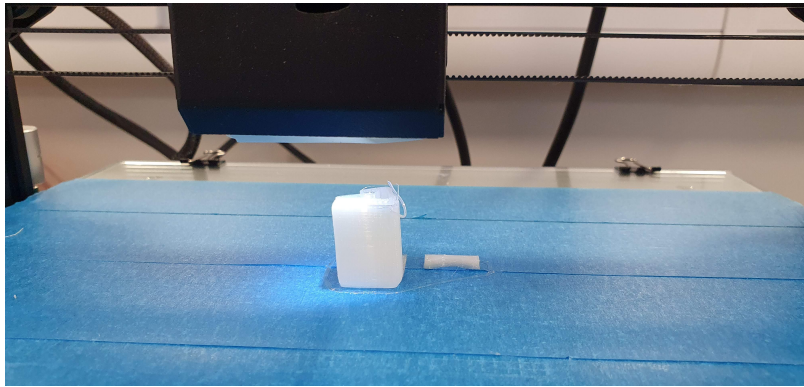


Figure 3.8: A 3D-printed sensor enclosure unit printed out of TPU using an FDM machine. The print was paused part-way through in order to embed the electronics inside the capsule.

The test mouthguards were submerged under warm water heated to 35°C, approximating intraoral temperature, and placed within a thermally insulated enclosure to ensure temperature stability for a duration of two hours. Following immersion, the mouthguards were thoroughly dried and subsequently cut open for inspection. The results found that all the additively manufactured units, all the way up to 3mm wall thickness, had failed the test, with water ingress evidenced by the moisture indicators papers turning from white to red. Upon inspection, this was likely due to two factors, the porous nature of additive-manufactured parts, and the sealing process of the cavity exhibiting imperfections. This means that it would be difficult to consistently

produce water-resistant mouthpieces through additive manufacturing. In contrast, all five thermoformed EVA units remained dry on the inside, with the moisture indicators papers remaining their original colour. Based on these findings, thermoformed EVA was selected as the preferred manufacturing method for subsequent iterations of the mouthguard, due to its superb reliability in preventing water ingress.

3.4.2 Battery Selection

One of the major risks identified during the risk assessment was potential battery damage due to heat exposure during mouthguard fabrication. EVA has an optimal thermoforming temperature in the range of $120^{\circ}\text{C} - 130^{\circ}\text{C}$ (Tanabe et al. 2020), with each forming cycle lasting approximately two minutes per layer. Accordingly, it is essential to select a battery capable of withstanding temperatures up to 130°C for durations of up to five minutes without degradation, in order to ensure both device integrity and user safety. An IEC62133-certified battery was therefore selected, as it is designed and tested to withstand exposure to temperatures up to 130°C for durations of up to 10 minutes without failure. This certification provides assurance that the battery can safely endure the thermoforming process without compromising its structural or electrochemical stability, thereby mitigating risks associated with thermal damage during fabrication. Ultimately, a Renata ICP501421PS-01 battery (Renata SA, a Swatch Group company, Itingen, Switzerland) was chosen based on its capacity, IEC 62133 certification, and the manufacturer’s established track record in supplying components for medical technology applications.

3.4.3 Ergonomic Design

One of the main objectives of this project is to design a sensor system that can monitor athletes unobtrusively on the field. In the initial pilot study, the instrumented mouthguard incorporated an electronics enclosure positioned on the palatal side be-

hind the teeth (see 3.2b). However, the participants reported noticeable discomfort resulting from contact between the device and the soft palate, which in some cases caused pharyngeal reflex. These observations prompted a redesign of the electronics placement to improve user comfort. Referencing existing smart mouthguards designed for head impact monitoring seen during the literature review (Kuo et al. 2018), a new design to host the electronic inside the outer wall of the mouthguard, with the battery placed inside the inner wall of the mouthguard to minimize the risk of direct impact, was proposed (see figure 3.9 for a CAD mock-up of the proposed design).

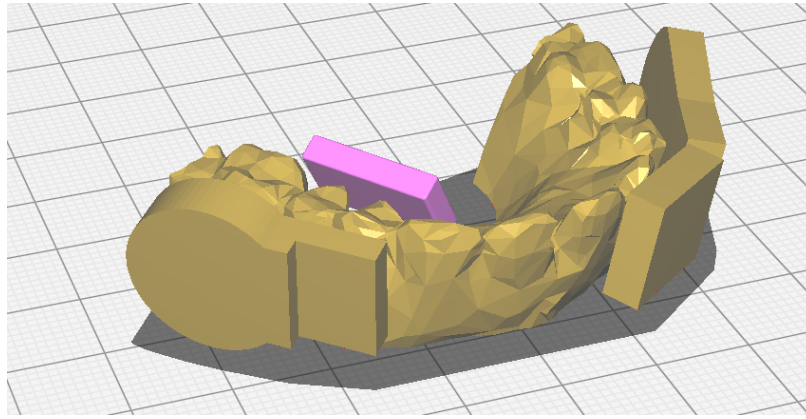


Figure 3.9: A 3D mock-up of the new board design concept, the pink block behind the teeth shows the proposed battery placement which minimizes the risk of mechanical impact on the battery.

To accommodate the integration of sensing components, onboard data storage, Bluetooth communication hardware, a wireless charging coil, and a battery within a custom-fitted mouthguard, a modular design strategy was adopted. In order to minimise device obtrusiveness and maintain anatomical conformity, a key design constraint was defined such that the maximum length of any individual electronics module should not exceed the combined mesiodistal width of the canine, two premolars, and one molar. This constraint enables the device to approximately conform to the natural curvature of the dental arch.

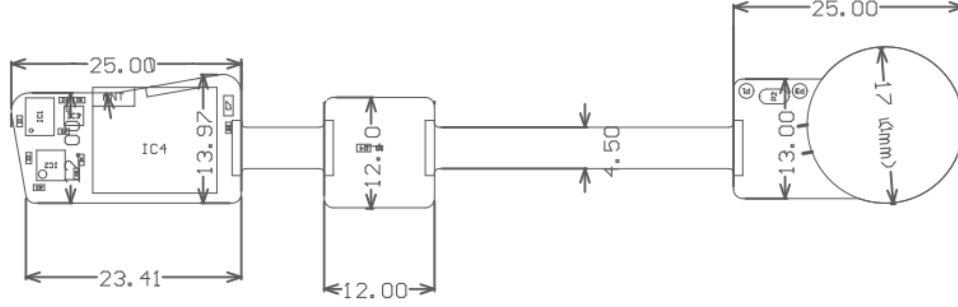


Figure 3.10: This figure depicts the finalized electronics design layout for the smart mouthguard: the left-most board contains the processor, microphone, IMU, storage, and Bluetooth device; the central board contains a PPG sensor for heart rate and blood-oxygen level sensing positioned above the lateral incisor (de Almeida e Bueno et al. 2023); the right-most board contains the power supply circuitry and a 17mm wireless charging coil.

Through out the project, 32 dental impressions were taken from 21 male and 11 female participants. It was noted that the dental impressions from the female participants are smaller than those of the male participants on average, which is inline with public dental datasets (Hashim & Sarah 2006). Using the public data described in Hashim & Sarah (2006), it was estimated that, to ensure compatibility with 95% of adult female user-base, the maximum allowable length of a single electronics module should be limited to 27mm. Working alongside internal and external electronics engineers who advised on the minimum required size for the modular boards based on electronics component-sizes, a final design with three modular boards was decided on. Figure 3.10 show a sketch of the final electronics layout.

3.4.3.1 Water Ingress Test Of The Updated Design

The revised board design was subjected to the same water ingress testing as previously conducted, in order to verify that the updated electronics placement within

the mouthguard did not compromise water tightness. Five mouthguards were made using mock-ups of the electronics, with a strip of moisture indicator paper running along the entire length of the mouthguard (see figure 3.11). They were submerged in a tank of water for two hours. Water ingress was observed in two of the five samples. Upon close inspection, it was determined that the water ingress was due to imperfect layer adhesion between the EVA sheets along the vertical axis.

The manufacturing process was thus revised with an additional step - a heated metal knife was used to melt and reshape the top edge of the mouthguards to ensure water tight layer adhesion. The water ingress test was carried out once more on the newly manufactured mouthguards, and found no evidence of water ingress in any of the samples. In addition, two mouthguards were subjected to prolonged immersion for 24 hours, with no detectable water ingress in either sample.

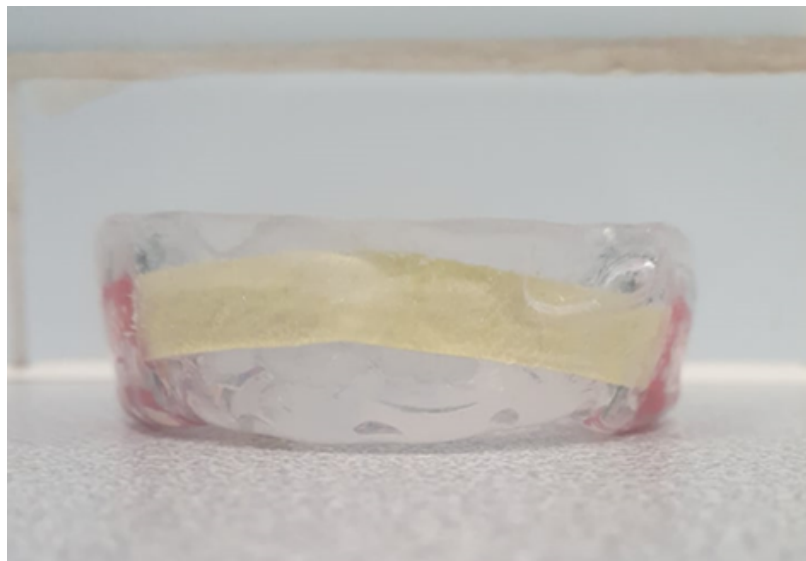


Figure 3.11: A custom mouthguard made from EVA via thermoforming, with the electronics embedded along the front wall of the mouthguard. A strip moisture indicator paper was embedded along the entire length of the mouthguard.

3.5 Mouthguard Manufacturing Process

This section outlines the finalized manufacturing process used to fabricate the instrumented mouthguards for the experiments.

First, a dental impression of the participant is obtained using a self-impression dental putty kit. Figure 3.12 shows the self-impression process. This process produces a negative cast of the teeth, which serves as the mould for the stone dental model.

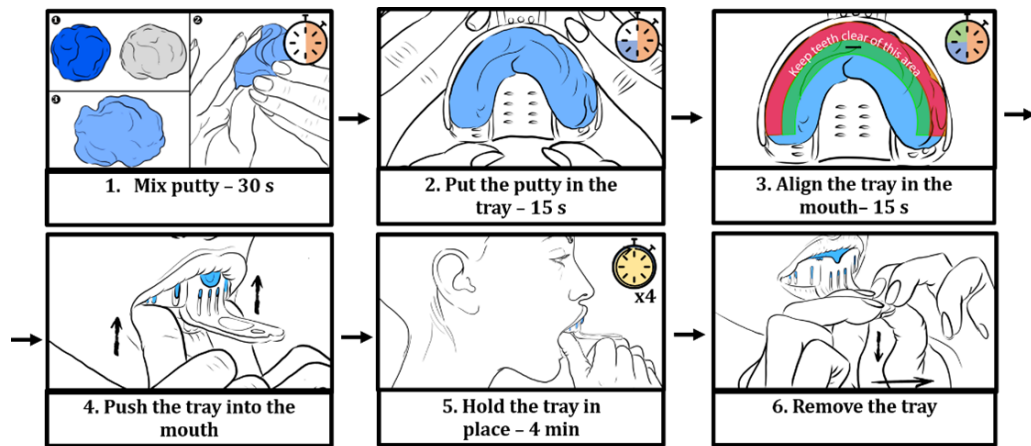


Figure 3.12: Summary of self-impression process

To create the stone dental model required for thermoforming the mouthguard, a dental-grade stone powder is mixed with water and gradually poured into the impression while placed on a vibration plate. The vibration helps eliminate air bubbles, ensuring a smooth and accurate model. Figure 3.13 shows an example of a finished stone dental model.

To fabricate the instrumented mouthguard, petroleum jelly is first applied to the stone dental model to prevent the EVA from sticking. A 1.5 mm EVA sheet is then heated at 130°C for 100 s and vacuum-formed over the model to create the base layer of the mouthguard.

While the base layer remains soft and malleable, the electronics are pressed into place, along the teeth, to secure them. A strip of moisture indicator paper is taped



Figure 3.13: An example of a stone dental model used to manufacture mouthguards

over the electronics as a safety feature, which will visually warn the user when fluid ingress has occurred. Excess EVA is trimmed, and the assembly is returned to the thermoforming machine.

A 3 mm EVA sheet is subsequently heated at 130°C for 180 s and vacuum-formed over the assembly, forming the outer layer of the mouthguard that encapsulates the electronics. Figure 3.14 shows the thermoforming process.

And finally, excess EVA is trimmed, and a heated metal knife is used to melt and reshape the top edge of the mouthguards to ensure water tight layer adhesion, completing the instrumented mouthguard (see figure 3.15).



Figure 3.14: This is an example of the thermoforming process. In this image, the electronics have already been assembled onto the mouthguard's base-layer, and the outer-layer is about to be vacuum-formed over the assembly.

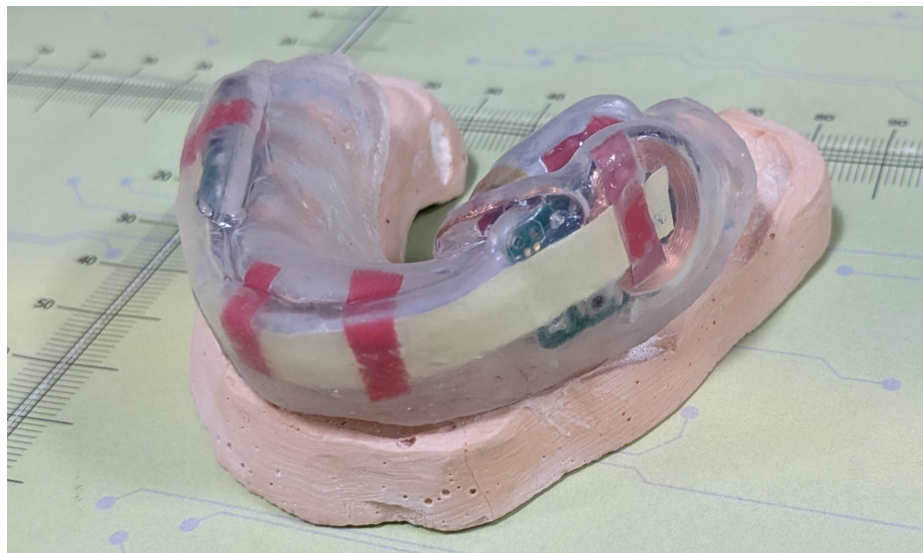


Figure 3.15: An example of the finalized version of the instrumented mouth-guard

3.6 Conclusion

This chapter examined the key design considerations involved in the development of a respiratory monitoring system for contact sports, focusing on sensor placement, user safety, and wearer comfort. It was shown that a thermoformed EVA mouthguard with integrated sensing can be unobtrusively deployed in contact sports. A series of pilot studies were conducted to evaluate the feasibility of the proposed device.

Pilot study results demonstrated that an intra-oral acoustic sensor can capture high-quality respiratory signals, enabling accurate respiratory rate estimation using a simple FFT-based algorithm under controlled conditions. These findings support the feasibility of a smart mouthguard as a respiratory monitoring platform.

However, limitations were identified in the performance of the FFT-based approach. Specifically, the algorithm exhibited reduced accuracy when applied to signals with irregular amplitude or variable breath duration, conditions that are likely to occur in real-world sporting scenarios. This highlights the need for more robust signal processing techniques capable of handling non-uniform and noisy data representative of the broader contact sports population.

Additionally, the observed correlation between breathing audio signals and minute ventilation (VE) suggests further potential for the proposed system to estimate more comprehensive respiratory metrics beyond breathing rate. To realize this capability, future work should focus on the collection of a larger, well-annotated dataset of breathing audio and the development of more advanced analytical methods.

Overall, this chapter establishes the feasibility of oral cavity acoustic sensing for respiratory monitoring and provides a foundation for continued system refinement, algorithm development, and validation under real-world sport conditions.

Chapter 4

Breathing Audio & Internal and External Workload Data Acquired During YoYo-IR1 Tests

Chapter Abstract

Last chapter presented a series of preliminary investigations showcasing the potential of an oral cavity based respiratory sensing wearable system. It was then concluded that a large breathing audio dataset with a wide range of participants needs to be collected during intense exercise, in order to further develop and evaluate the wearable solution. This chapter details the major data collection work which generated the dataset used to support the studies in the subsequent chapters of this thesis. In this study, 21 healthy participants of different age, sex, and level of fitness were recruited to capture a wide spectrum of human physiological differences. The participants were instructed to undertake Level 1 Yo-Yo Intermittent Recover Tests (Yo-Yo IR1 tests), while equipped with an instrumented mouthguard (see Appendix ??) to capture continuous breathing audio, and a COSMED K5 portable metabolic cart to capture the

ground truth of respiratory metrics such as respiration rate and minute ventilation, as well as other workload measures such as heart rate (though a Polar H10 paired with the COSMED K5) and running speed (via onboard GPS). During the tests, the participants were instructed to indicate their rating of perceived exertion (RPE) each time they crossed the start line of the test, using RPE levels written out in chalk on the start line.

Additionally, this chapter presents tests conducted on the sensing equipment to ensure data completeness. The microphone board used in the instrumented mouth-guard was found to have a time drift of $< 280ppm$ which is accounted for by the inaccuracy of the onboard real-time-clock. The COSMED K5 was found to have an auto-recalibration period of around 10s for every 5min of data recording.

Chapter Abstract

This chapter presents the primary data collection effort underpinning the subsequent analyses in this thesis. The aim of this study is to develop a comprehensive dataset of breathing audio and associated internal and external workload metrics collected during high-intensity intermittent exercise, in order to enable the evaluation of a wearable respiratory sensing system.

A cohort of 21 healthy participants with diverse age, sex, and fitness levels were recruited to capture a broad range of physiological responses. Participants performed Level 1 Yo-Yo Intermittent Recovery Tests (Yo-Yo IR1) while equipped with an instrumented mouthguard to record continuous breathing audio. Simultaneously, a COSMED K5 portable metabolic cart was used to obtain ground-truth respiratory and physiological measurements, including respiration rate, minute ventilation, heart rate, and running speed. Subjective workload was also captured using ratings of perceived exertion (RPE) reported throughout the test.

The resulting dataset provides synchronised multimodal measurements of breathing acoustics, physiological responses, and physical workload during progressive high-intensity exercise. Equipment validation tests were also conducted, identifying and quantifying sources of measurement drift and calibration interruptions.

This dataset forms the foundation for the modelling, analysis, and validation work presented in Chapters 5–7. Beyond this thesis, the dataset contributes a novel resource for investigating the relationship between breathing acoustics and physiological workload, with potential applications in wearable sensing, sports performance monitoring, and fatigue assessment.

4.1 Introduction

Top-level athletes have long used physical and physiological performance monitoring and analytics in training to gain a competitive advantage. The most notable example of such practices is through the use of metabolic carts, such as the Parvo Medics TrueOne 2400 (ParvoMedics, Inc, Salt Lake City, Utah, USA), to measure cardiopulmonary measures such as VO_2max , and determine fitness improvements over time. These metabolic tests involve extremely expensive and often immobile equipment and can only be carried out in a lab environment with the help of well-trained technicians (Buttar et al. 2019). Due to these restrictions, physical and physiological performance monitoring had been inaccessible to most in the past, however, this has changed in more recent times with new advances in wearable technology and data science (Muniz-Pardos et al. 2021, Li et al. 2016). Wearable sensor systems have facilitated the collection of a wide range of physical and physiological information on the field in team sports, ushering in a new age of sports monitoring and analytics.

One major emerging trend in sports monitoring is fatigue monitoring. And while there exists a disparity in the literature regarding the precise usage and definition of the word fatigue (Jones et al. 2017), the consensus in sports science is that the state of fatigue is indicated by a decrease in one’s physical and mental capacities to the point of impairing one’s usual abilities (Jones et al. 2017, Thorpe et al. 2017, Halson 2014, Abbiss & Laursen 2005). The objectives of fatigue monitoring include injury and illness forecast and prevention, informing return-to-play decisions for injured players, accessing the effectiveness of training routines, optimizing match outcomes in team sports, and more (Jones et al. 2017, Thorpe et al. 2017, de Almeida e Bueno et al. 2021, Halson 2014, Cheng & Bergmann 2022). There isn’t a consensus on how to measure fatigue in the literature, though, most hypothesize that fatigue needs to be measured holistically utilizing off-field factors such as sleep; and on-field measure-

ments of physical workloads (e.g. distance covered), referred to as external loads in sports science, as well as the physiological responses (e.g. heart rate) to the workloads, referred to as internal loads (Jones et al. 2017, Thorpe et al. 2017, Halson 2014, Cheng & Bergmann 2022).

The measurement of external workloads is relatively straightforward, utilizing kinematic measurements derived from technology such as global positioning systems (GPS), local positioning systems (LPS), and inertial measurement unit-based (IMU) positioning systems (Cheng & Bergmann 2022, Cardinale & Varley 2017, Li et al. 2016). Measuring internal workloads, on the other hand, is a more multifaceted problem, with a large array of potential metrics including vitals such as heart rate (HR) and respiratory frequency, blood and muscle oxygen saturation (SpO2 and SmO2), hormonal bio-markers, and rating of perceived exertion (RPE), just to name a few. With a wide range of options available, there is very little effort to investigate the viability of measuring fatigue using multiple internal workload metrics in tandem (Cheng & Bergmann 2022). Additionally, due to accessibility to available technology, there is a stronger focus, in the literature, on easy-to-assess metrics like heart rate (and derivative metrics) and RPE than respiratory metrics which are less accessible (Cheng & Bergmann 2022, Benson et al. 2020, Seshadri, Li, Voos, Rowbottom, Alfes, Zorman & Drummond 2019, Nicolò et al. 2017, Nicolò, Massaroni, Schena & Sacchetti 2020). Limited studies have shown respiratory audio information as promising internal workload metrics, with respiratory audio amplitude demonstrated as a possible stand-in for tidal volume (de Almeida e Bueno et al. 2021), and analysis of breathing audio as a plausible surrogate for the gas exchange threshold (Peterson et al. 2014); while respiratory frequency exhibits good agreement with RPE in standardized fitness tests (de Almeida e Bueno et al. 2021, Nicolò, Massaroni, Schena & Sacchetti 2020). This warrants a further investigation into incorporating respiratory information when monitoring fatigue in team-sports athletes.

To facilitate the exploration of relationships between different external and internal workloads, and the development and validation of ways to measure fatigue using multiple external and internal workload metrics, this study was conceived. In this study, 21 healthy human participants of various fitness levels were recruited to partake in a standardized running interval test (The Yo-Yo Intermittent Recovery Test Level 1) (Bangsbo et al. 2008). During the interval test, each participant is equipped with a COSMED K5 (COSMED, Rome, Italy) portable metabolic machine, a Polar H10 (Polar Electro Oy, Kempele, Finland) heart rate monitor, and a microphone-instrumented mouthguard. A modified RPE scale was integrated into the running test course to track RPE changes throughout the test (de Almeida e Bueno et al. 2021). The published dataset consists of pre-processed breathing audio data, breath-by-breath data output from COSMED K5 (RF, VT, VE, VO₂, VCO₂, HR, EE, etc.), Yo-Yo Intermittent Recovery Test (Yo-Yo test) levels with time stamps, and RPE scores corresponding to the Yo-Yo Intermittent Recovery Test levels.

Using this dataset, the subsequent chapters of the thesis addressed the feasibility of monitoring respiratory rate and minute ventilation, as well as estimating RPE level using the microphone-instrumented mouthguard system described in Chapter 3.

4.2 Methods

4.2.1 Recruitment

This study was conducted with the approval of the Research Ethics Committee of the University of Oxford (R43470/RE001). For the study, 21 participants, between the ages of 18 to 43, were recruited. The participants were all well-informed on the purpose of the study and have given written consent to be included in the publications resulting from this work. All participants included are self-declared to be healthy and in shape to carry out the physical activities detailed in the experimental protocol.

All participants reside in the UK and are proficient in English. The participants were invited to two identical data capture sessions scheduled weeks apart, all 21 participants participated in the first round of data captures, and 19 participants returned for the second round of data captures. This was done to capture variability within each participant and minimize the potential bias due to environmental factors, as well as to generate more data per participant for ML training applications.

Basic information about each participant was collected at the beginning of each data capture session, which includes age, gender, and self-reported weight and height. Of the 21 participants, 10 were female and were between the ages of 19 to 40, with a weight distribution ranging between 50kg and 87kg, and a height distribution ranging between 150cm and 182cm. The other 11 participants were male and were between the ages of 18 to 43, with a weight distribution ranging between 63kg and 92kg, and a height distribution ranging between 170cm and 194cm.

4.2.2 Experimental Setup

In this study a multi-stage running fitness test was used to simulate external workloads experienced by athletes in field sports, and the internal and external workloads were captured using various wearable sensor systems. In this study, GPS speed tracking, Borg’s Category Ratio 10 scale (CR10) ratings of perceived exertion (RPE) (Borg 1990), and cardiopulmonary and metabolic measurements were collected from the participants during Yo-Yo intermittent recovery level 1 (Yo-Yo IR1) tests (Bangsbo et al. 2008).

4.2.2.1 Equipment

To capture internal and external workloads during the Yo-Yo IR1 tests, the participants wore a COSMED K5 mobile metabolic cart, which captures a range of respiratory and metabolic measurements as well as GPS information. The COSMED

K5 was used in the breath-by-breath (BxB) data collection mode for the purpose of this study. The COSMED K5 was calibrated using a standard reference gas canister at the beginning of each data collection day, and the flowmeter on the COSMED K5 was re-calibrated before each data capture session. Heart rate data was collected via a Polar H10 heart rate monitor strap that was wirelessly connected to the COSMED K5.

Some participants also consented to wearing a microphone-instrumented mouthguard (as described in Chapter 3), which collected audio signals of respiration. Nineteen participants consented to wearing an instrumented mouthguard in the first round of data collection, and thirteen participants consented to wearing it in the second round of data collection. To custom-fit the instrumented mouthguards to the participants, dental impressions of the participants' upper teeth were collected during the experimental protocol briefing session via dental self-impression kits. The mouthguards were then thermoformed out of dental-grade EVA to the dental impressions. The encapsulated microphone device was an MP34DT05-A digital microphone (STMicroelectronics, Switzerland), with a 64 dB signal-to-noise ratio and $-26dBFS \pm 3dB$ sensitivity.

4.2.2.2 Ratings of Perceived Exertion

One of the most widely applied methods of evaluating levels of fatigue in sports is the ratings of perceived exertion (RPE) (Cheng & Bergmann 2022, Faulkner et al. 2008). Since its first introduction, the RPE scale has gone through many iterations and spawned numerous variants (Faulkner et al. 2008). The basic idea behind RPE scales is that numerical values are assigned to subjective intensities of physical activities, the scales follow a ratio scaling pattern, and by using well-defined anchor phrases to describe key points on the scales, perceived exertion can be estimated with a level of consistency across different individuals (Borg 1990). A modified version of the Borg's

CR10 RPE scale was used in this study (Borg 1990, Foster et al. 2001, de Almeida e Bueno et al. 2021) for its ease of use and comprehensibility, with a straightforward 0 to 10 scaling, and clear and simple anchor words. The modified RPE scale for this study is shown in table 4.1.

Rating	Descriptor
0	Rest
1	Very Easy
2	Easy
3	Moderate
4	Somewhat Strong
5	Strong
6	*
7	Very Strong
8	*
9	*
10	Extremely Strong

Table 4.1: Modified RPE scale used in this study. The participants were shown the scale before each data capture session. The scale along with the descriptors was also written in chalk on the 0m line of the running track. During the Yo-Yo IR1 tests, the participants were instructed to indicate their RPE level by stepping on the corresponding number each time they crossed the 0m line. The session RPE was also collected from the participants approximately 20 minutes after each data collection session.

4.2.2.3 Test Protocol

Most team sports involve intermittent running exercise, so to simulate the physical demands of such sports, the Yo-Yo IR1 test (Bangsbo et al. 2008) was selected for this study. The Yo-Yo IR1 test has been extensively used in testing the fitness of athletes in a number of team sports, such as soccer (Rampinini et al. 2010, Krstrup et al. 2006, Castagna et al. 2006, Krstrup et al. 2003), rugby (Atkins 2006, Higham et al. 2013), basketball (Castagna et al. 2008), and handball (Souhail et al. 2010), and

has been found to be a representative assessment of aerobic fitness and game-related endurance for these sports (Bangsbo et al. 2008, Krstrup et al. 2003, Atkins 2006, Higham et al. 2013, Castagna et al. 2008, Souhail et al. 2010). The test consists of bouts of progressive speed shuttle-running around a 20m track with 10 seconds of active recovery, in the form of a short walk or jog around a 5m track behind the start line, in between. The running speed is paced via audio cues, and the test terminates when two consecutive cues have been missed. Figure 4.1 shows the layout of the test track.

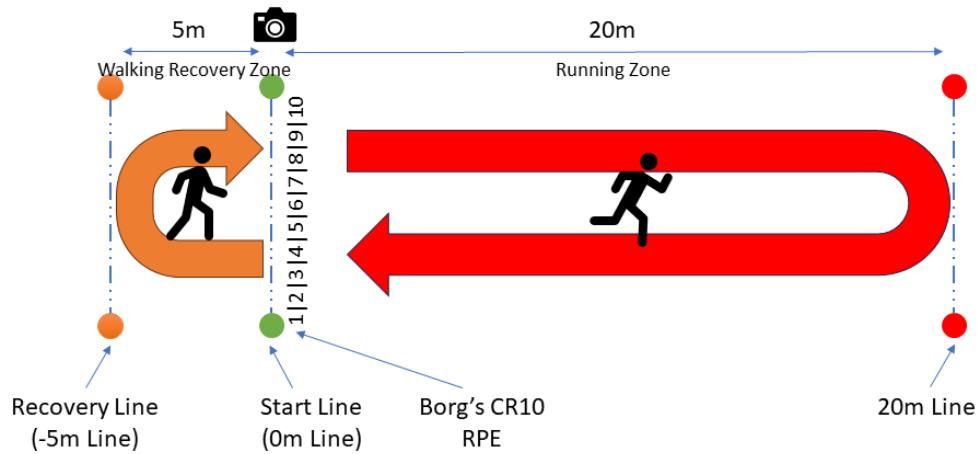


Figure 4.1: This figure shows the setup for the Yo-Yo IR1 test. The track consists of two zones, a running zone spanning 20m across, and a walking recovery zone spanning 5m. The Borg CR10 RPE scale is marked in front of the 0m line. A researcher is placed on the 0m line, recording the participant's RPE level during the test, with pen and paper and on a camera.

A hard-surface outdoor multipurpose sports court was used to carry out the Yo-Yo IR1 tests in this study. At the start of every data capture session, the participants were shown a card explaining the modified RPE scale and given time to familiarize themselves with it before the tests began. They were then equipped with the sensor systems described above (COSMED K5 portable metabolic cart, Polar H10 heart rate

monitor, and microphone-instrumented custom mouthguard). A mandatory warmup then took place, where the participants were instructed to stretch and then take the first two levels of the Yo-Yo test, which ensured they understood the test. When the participants were ready, if they had consented to wearing an instrumented mouthguard during the test, they were told to stand on the 0m line on the track and perform an audio-signal marking task, where they held their breath for 10s and exhaled while saying 'ah' to mark the start of the audio data collection. The participants were also instructed to remove the instrumented mouthguard right after the Yo-Yo test terminated.

To capture the participants' RPE level, the modified CR10 RPE scale, along with the descriptors, was written in chalk on the 0m line of the running track (see figure 4.2). During the Yo-Yo IR1 tests, the participants were instructed to indicate their RPE level by stepping on the corresponding number each time they crossed the 0m line. A researcher was stationed at the 0m line recording the run on video. This was done to both record the RPE value as well to help align the audio and COSMED data at a later stage. Figure 4.3 shows a snapshot of the experimental setup at the 0m line of the Yo-Yo test, with the participant fully equipped with the sensor systems. The session RPE (Foster et al. 2001), rated on the same scale, was also collected from the participants at the end of each data collection session after a recovery period.

4.2.3 Preprocessing

The Polar H10 has, in some rare cases, malfunctioned during this study. Although the exact cause of the Polar H10 malfunctioning is uncertain, this author believes it was likely due to a low battery level. Heart rate data resulting from the Polar H10 malfunctioning have been omitted from the final version of the dataset, a column in the metadata files shows the availability of heart rate data for each participant. Additionally, motion artifacts resulting in the heart rating measuring as 0 can be



Figure 4.2: This figure shows the Borg CR10 RPE scale, written in chalk, marked in front of the start (0m) line. The participants were instructed to step on the number corresponding to their RPE level as they cross the start line. Behind the start line, cones are placed 5m away to indicate the recovery zone of the YoYo test.

seen in the data. While these artifacts were not removed from the final version of the dataset, code to fill in the gaps based on the nearest non-zero values was included in the data repository.

COSMED K5 data are exported as *.xlsx files initially, containing information about the participants, environmental measurements (temperature, relative humidity, and air pressure) during the data capture sessions, and physiological measurements recorded breath-by-breath (BxB) during the sessions. The personal information and environmental measurements are transferred over to a metadata file and removed



Figure 4.3: This figure shows a participant, wearing the COSMED K5, a Polar H10 heart rate strap, and a microphone-instrumented mouthguard, at the starting (0m) line at the beginning of a Yo-Yo test level. The Borg CR10 RPE scale, written in chalk, was marked in front of the start (0m) line. The participant can be seen stepping on the number 3, indicating their RPE level was 3 at that instant.

from the COSMED data files, and the remaining physiological measurements are then converted into *.csv files. To identify the start time of the YoYo test in the COSMED data, the speed data captured by the onboard GPS was analyzed. Since the participants were instructed to stand still on the 0m line of the track before the start of the test, the test start time can be manually annotated on a time series plot as show by figure 4.4.

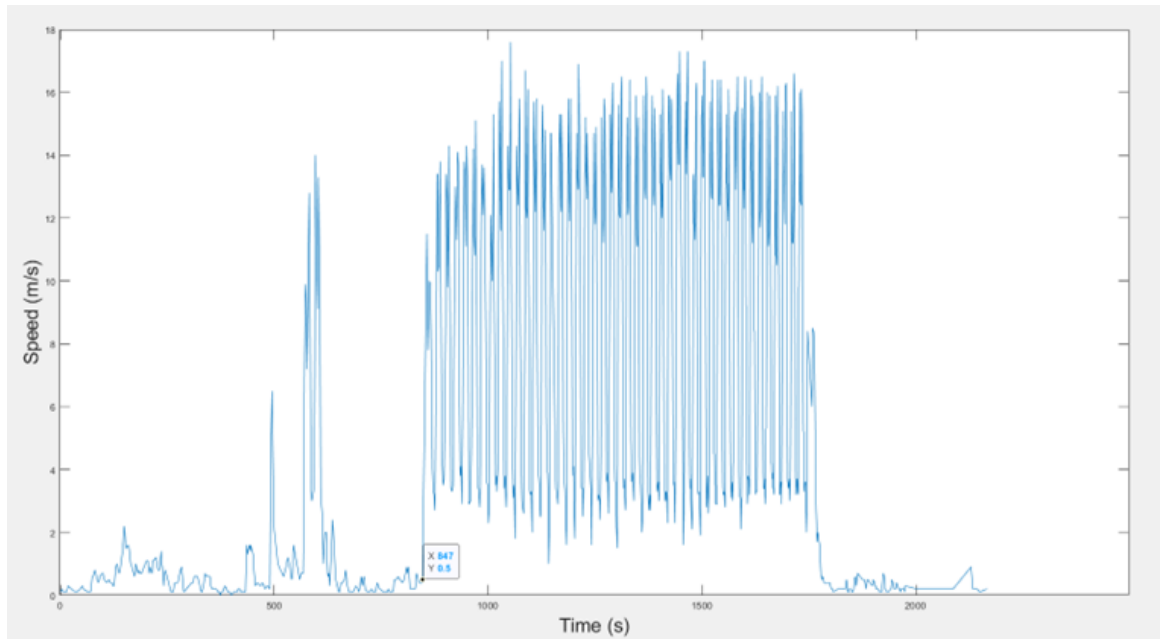


Figure 4.4: This figure shows an example of the GPS speed data captured by the COSMED. In this example, the start of the YoYo test has been highlighted.

To identify the start time of the YoYo test in the audio data, the audio-signal marking task was manually annotated for the audio data. The time between each audio-signal marking task and its corresponding Yo-Yo test start time were measured to the nearest second using video playback (VLC media player, VideoLAN, Paris, France) in the video recordings taken at the 0m line. This was then used to identify the start time of each Yo-Yo test in the audio data.

Since the COSMED data is recorded on a breath-by-breath basis, the audio data

and the COSMED data can still be misaligned by up-to the duration of a full breath. To further align the two, breath-by-breath breathing rate was algorithmically measured from the audio data (see chapter 5 for details), and cross-correlation was used to finetune the alignment between the two data types.

4.3 Technical Validation

Technical validation for both the Polar H10 and the COSMED K5 exist within the literature (Gilgen-Ammann et al. 2019, Schaffarczyk et al. 2022, Perez-Suarez et al. 2018, Winkert et al. 2020, Crouter et al. 2019). The Polar H10 has been shown to be capable of capturing heart rate data accurately during low to high-intensity exercises (Gilgen-Ammann et al. 2019, Schaffarczyk et al. 2022); the COSMED K5 has been shown to be capable of capturing physiological data accurately at low to moderate metabolic rates, while underestimating oxygen consumption (VO_2) at high metabolic rates (Winkert et al. 2020). The microphone-instrumented mouthguards used in this study were presented in previous studies (de Almeida e Bueno et al. 2021, Cheng & Bergmann 2023), where they have been shown to be capable of collecting respiratory information during high-intensity running exercise and while under noisy outdoor conditions.

4.3.1 Microphone Board Data Completeness Test

This test was designed to verify the microphone board, used in the instrumented mouthguards to capture breathing audio, is capable of recording data fully and correctly over the duration of the experiment described above.

4.3.1.1 Method

A sinusoidal signal of 200Hz modulated by a sinusoidal wave of 1Hz was encoded into a wav audio file with a sampling frequency of 40kHz. The audio file is played on a loud speaker continuously for exactly one hour in a quiet room with a background noise level lower than 20dB. The microphone board, encapsulated in EVA (the material used to manufacture the instrumented mouthguards) while connected to a stable power source, was placed next to the speaker and left recording for the full duration of the audio file playback. Afterward, the open-source audio data annotation software Praat (Boersma 2001) was used to analyze the recording in both time and frequency domains.

4.3.1.2 Results

Initially, a mismatch between the reference signal and the recorded signal was found, and data loss was detected at random intervals. The firmware of the system was subsequently updated to address this issue, and the system was once again tested for data completeness. For the test after the firmware update, the frequency domain analysis shows that the recorded signal has a carrier frequency of 200Hz and pulses once per second, which is inline with the reference signal. The time domain analysis shows that the hour-long recording clocking at just slightly under 3601 seconds, meaning there is a one second time drift over the course of an hour. This equates a time error of around 280 parts per million (ppm). This error is on the same scale as the accuracy listed for the real-time-clock on both the microphone board and the device used to play the test signal. Thus, it was concluded that the microphone board encapsulated in EVA is capable of correctly capturing breathing audio during running fitness tests such as the Yo-Yo IR1 test that last for less than half an hour.

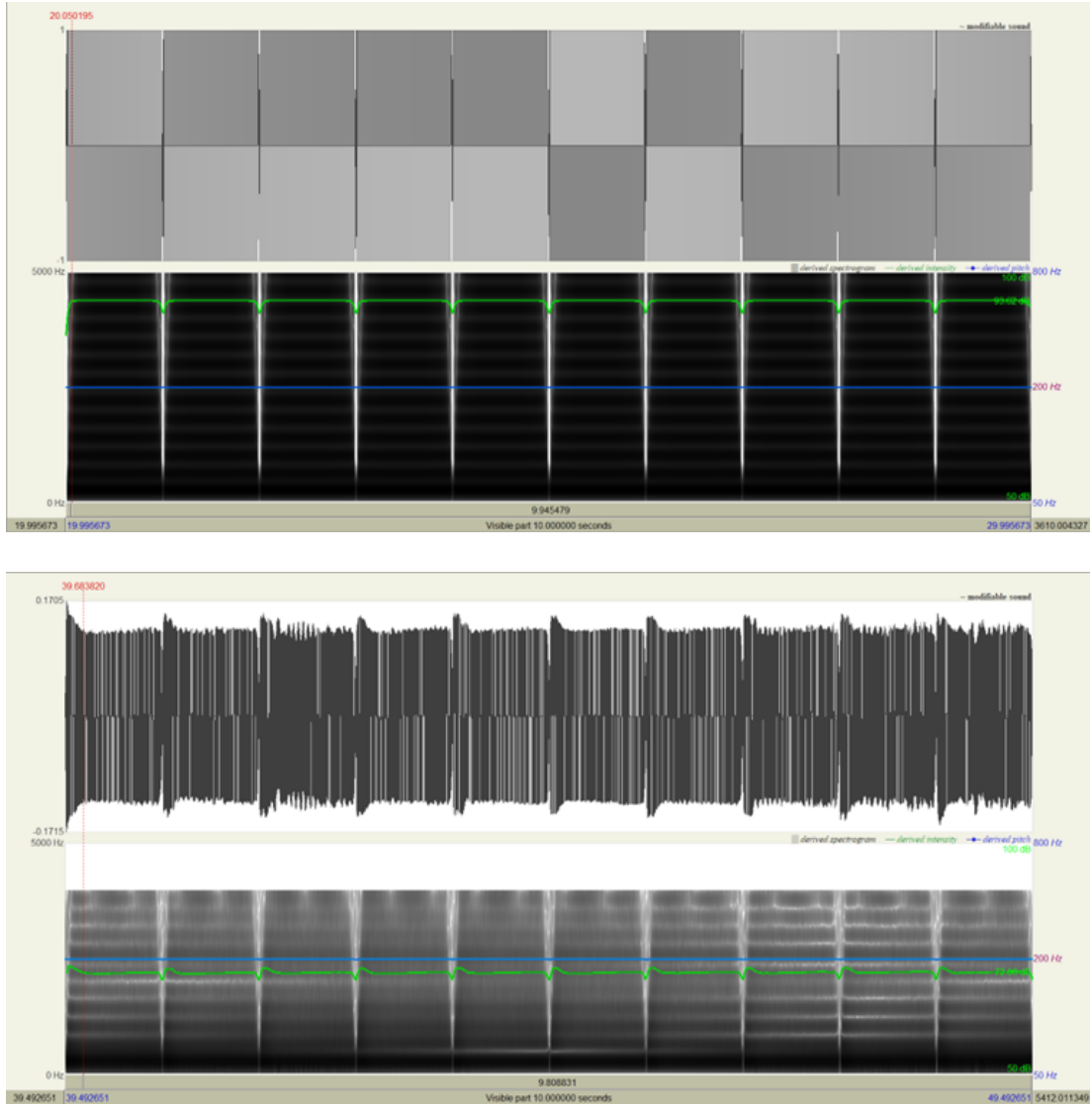


Figure 4.5: Top: a 10s clip of the computer generated test signal, 10 pulses can be seen in the time domain and the frequency domain shows the carrier wave is at 200Hz. Bottom: a 10s clip of the test signal recorded by the microphone board, it exhibits a similar pattern to the original signal shown above.

4.3.2 COSMED K5 Data Completeness Test

During data preprocessing, it was noticed that the COSMED has exhibited a pattern of reoccurring short-interval data loss. A deep dive into the K5 user manual found that

there's a single reference to an auto-calibration cycle, but there is no documentation on how this calibration mechanism impacts data collection. It was suspected that the increase in accuracy of gas analysis occurred at the expense of coverage of breaths during measurements. Meaning that while the system was re-calibrating, breath measurements were omitted. This period of un-measured data, however, is not clearly labelled on the system output. Thus, a further lab-based experiment was carried out in an attempt to characterize the data loss pattern.

4.3.2.1 Data Acquisition

The lab-based experiment consisted of four tests, completed by one subject over two consecutive days. For each test, the subject was sat upright in a stationary chair in a resting state. Once the subject was connected to the COSMED K5 system, the experiment began by allowing the system to automatically begin recording data. This occurs two minutes after the first breath is registered in the 'pre-test phase'. Once recording began, the subject was required to breathe at a constant frequency for twenty minutes, which was paced by a metronome. The subject also wore a microphone-instrumented mouthguard to act as an additional reference for breathing frequency during the measurement duration. The test was completed for two frequencies, 15 and 30 breaths per minute, with each test undertaken once per day.

4.3.2.2 Analysis

Data loss is noted when visually examining time-domain parameters such as timestamps (t) and breath duration (T_{tot}). Table 4.2 shows an example of BxB (at a constant breathing rate) data output from COSMED K5 when exported to a spreadsheet. The difference between the timestamps and the cumulative T_{tot} (denoted as δ) should remain relatively constant over time when data loss is absent, thus, by plotting δ and its step change (denoted as δ_{diff}), the data loss can be visualized and

	t(s)	Ttot(s)	tc(s)	tdiff(s)
Breath 1	00:14	4.05	4.05	–
Breath 2	00:18	3.98	8.03	4
Breath 3	00:22	4.07	12.10	4
Breath 4	00:34	4.11	16.21	12
Breath 5	00:38	3.98	20.19	4

Table 4.2: Table showing an example of BxB data output from COSMED K5. The first column of the table is annotation data output for five breaths. The second column shows t, the timestamps of the data output. The third column shows Ttot, the total breath duration. The fourth column shows tc, the cumulative sum of total breath durations. The fifth column shows tdiff, the difference between successive timestamps. The red line indicates where t significantly deviates from tc which highlights where there might be an issue of data loss.

quantized (figure 4.6).

4.3.2.3 Results

It was found that δdiff ranged from 8 to 12 seconds during data loss segments, averaging to around 10 seconds each time. It was also found that the first data loss segment in each test occurred at exactly 170 seconds, factoring in the 2-minute 'pre-test phase', the data loss consistently occurred at 4 minutes 50 seconds after the start of the COSMED measurement, while lasting around 10 seconds. Moreover, it is noted that each consequence data loss occurred 5min apart. Thus, it is concluded that the data loss due to automatic re-calibration is systematic, occurring every 5 minutes for around 10 seconds. Based on this finding, all COSMED data files captured in this study were analyzed and all periods of missing data were identified.

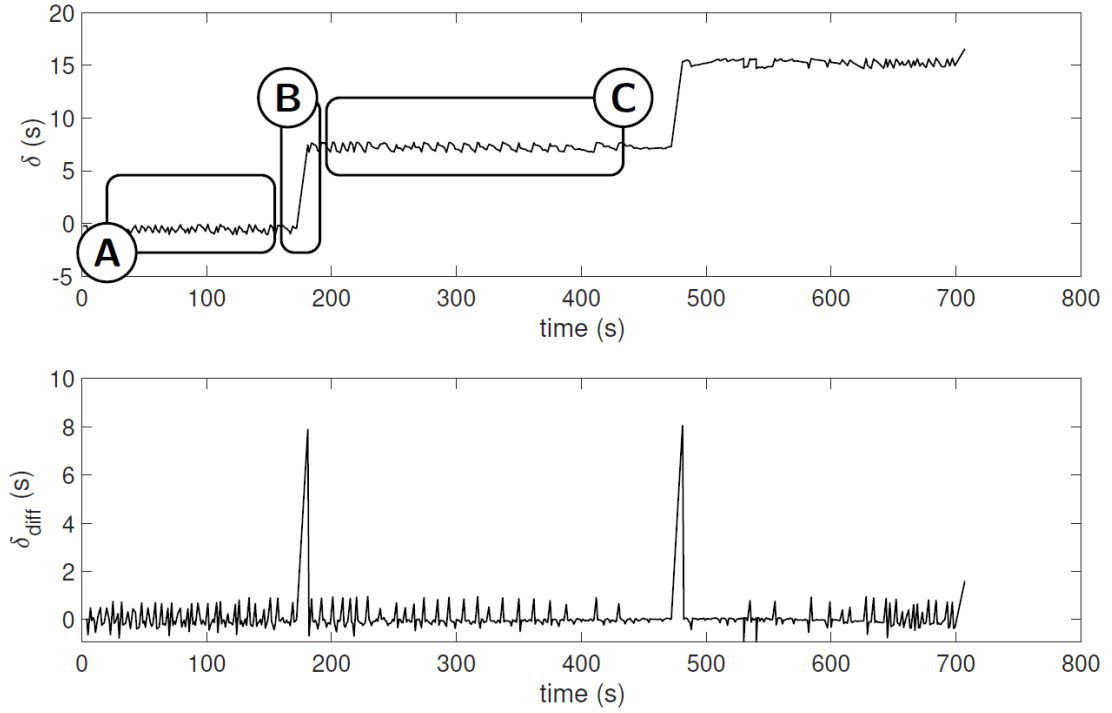


Figure 4.6: Top: Region A highlights normal recording activity where δ is constant and close to zero at the beginning of the recording. Region B highlights a period of unregistered breaths where the real-time clock responsible for t continues to run while t_c stops, resulting in an increasing δ . Region C highlights the return to normal recording activity where δ settles at a new constant. Bottom: The size of the step-change in δ is approximately equal to the duration of the data loss, this is represented by δ_{diff} . (Figure credit to William R. F. Milnthorpe, whose contribution was essential to this investigation.)

4.4 Usage Notes

This dataset presents a variety of potential applications relevant to sports science and athlete monitoring. One key use involves analyzing the strength of correlations between internal and external workload metrics during a multistage running test, offering insights into the physiological demands of exercise. It also supports the devel-

opment and validation of an audio-based, breath-by-breath respiratory rate detection algorithm specifically designed for use in sports contexts. In addition to respiratory rate, the dataset enables further exploration of respiratory audio signals to extract other internal workload metrics, such as minute ventilation and tidal volume, potentially enhancing the depth of on-field physiological assessment. Furthermore, the dataset may facilitate the development of composite workload indicators by exploring how different internal and external metrics can be integrated to more accurately reflect fatigue.

4.5 Data Records

The dataset was uploaded to GitHub at <https://github.com/Oxford-NIL/YoYoTest>. The dataset was split into two subsets for the two rounds of data collection. At the top level, there are two sub-folders each containing one of the subsets. There's also a readme file with a brief description of the experiment and data structure. Also included at this level, are two scripts, one in Python and one in MATLAB, for loading the dataset.

Inside each subset folder, there is a metadata file detailing the data included within the subset, from participants' information, to environmental measurements during the data capture sessions, to details about the sensor data collected from each participant (See, in Appendix C, tables C.1 C.2 C.3 for details about subset one, and tables C.4 C.5 C.6 for details about subset two). You will also find individual folders for each participant under each subset folder, inside of which contain three unique files, a *.wav file for the respiratory audio data, a *.csv file for the COSMED K5 data, and a *.csv for the RPE levels recorded at the start of each Yo-Yo test level with timestamps.

The COSMED data files contain more than 50 different measures, for the full

breakdown, please refer to COSMED data files in the dataset. The most commonly used measures (in sports) contained in the files include - respiratory frequency (Rf) along with inhale duration (Ti) and exhale duration (Te), tidal volume (VT), minute ventilation (VE), oxygen uptake (VO₂), carbon dioxide production (VCO₂), heart rate (HR), energy expenditure (EE), and GPS derived speed.

4.5.1 Recorded Modalities

The dataset consists of the following synchronized data streams:

- **Breathing audio:** Captured continuously using an instrumented mouthguard.
- **Metabolic metrics:** Including respiration rate and minute ventilation, measured using the COSMED K5.
- **Cardiovascular data:** Heart rate measured via Polar H10.
- **External workload:** Running speed derived from COSMED K5 onboard GPS.
- **Subjective internal workload:** Ratings of perceived exertion (RPE).

4.6 Conclusion

This chapter presented the primary data collection effort for this thesis, resulting in a comprehensive multimodal dataset of breathing audio, physiological measurements, and GPS data acquired during Yo-Yo IR1 tests.

The dataset captures a wide range of physiological responses across participants with varying characteristics, providing a robust foundation for investigating the relationship between breathing acoustics and internal workload. Equipment validation ensured the reliability and interpretability of the collected data.

This dataset is used extensively in the following chapter of the thesis to develop and evaluate models for estimating physiological workload from breathing audio and

to explore its potential for fatigue monitoring in sports contexts.

Beyond this thesis, the dataset represents a valuable resource for the broader research community. It enables further investigation into breathing audio processing, respiratory sensing, and multimodal performance monitoring. Plans for future work include preparing the dataset as well as the experiment methodology for publishing on PhysioNet, to support reproducibility and facilitate further research in this domain.

Chapter 5

Breathing Event Detection From Audio Signal

Chapter Abstract

Chapter Abstract

Building on the demonstration in Chapter 3 that high-quality respiratory audio can be acquired from the oral cavity, and on the outdoor exercise dataset developed in Chapter 4, this chapter evaluates whether a microphone-instrumented mouthguard can be used to detect breathing events and estimate breathing rate during high-intensity intermittent exercise.

Respiratory audio recorded during Yo-Yo intermittent recovery tests was processed by transforming the signal into mel spectrograms and averaging across frequency bins to obtain a respiratory sound envelope. A trough-detection algorithm was then applied to identify respiratory flow reversals, corresponding to the onset of inspiration or expiration. The algorithm was evaluated against 3555 manually annotated res-

piratory flow reversals and against breathing-rate measurements from a COSMED K5 portable metabolic system. For the COSMED comparison, both audio-derived and reference breathing-rate signals were interpolated, smoothed using a 10s moving-average filter, aligned using cross-correlation, and assessed using mean absolute error, mean absolute percentage error, mean of differences, and limits of agreement.

Against manual annotations, the algorithm achieved high event-detection accuracy, with a precision of 98.8%, recall of 98.1%, and a mean absolute timing error of 0.03s. Compared with the COSMED K5, the audio-derived breathing rate showed close agreement, with an $MoD \pm LoA$ of -0.1 ± 9.3 bpm, an MAE of 3.0bpm, and an MAPE of 7.2%. The findings are comparable to similar studies measuring breathing rate using microphones in other wearable configurations. These findings show that respiratory audio captured by a mouthguard-mounted microphone can enable accurate detection of breathing events and practical estimation of breathing rate during high-intensity outdoor exercise.

The key contribution of this chapter is the validation of an interpretable, spectrogram-envelope-based algorithm for respiratory monitoring using an instrumented mouthguard in realistic field conditions. This supports the broader thesis that mouthguard-based acoustic sensing is a viable approach for on-field respiratory monitoring in sport, particularly in settings where conventional wearable systems may be limited by motion artifacts, comfort, or suitability for contact-sport use.

5.1 Introduction

Breathing rate (BR) is an important physiological measure that can provide valuable insights into a person’s health and fitness. In medicine, BR has been linked to cardiovascular health and shown to be an important predictor of cardiac arrest (Cretikos et al. 2008). In sports physiology, BR has been shown to be closely linked to psychophysiological metrics like the rating of perceived exertion (RPE) (de Almeida e Bueno et al. 2021, Nicolò et al. 2016). If leveraged appropriately, continuous monitoring of BR as an internal workload metric during training can help prevent overtraining and even injuries.

Several wearable sensor system options for capturing breathing rate have been identified in the literature since the start of this project. Mainly, four categories of technology were observed (Massaroni et al. 2019, Liu et al. 2019, Costanzo et al. 2020, Vanegas et al. 2020, Vitazkova et al. 2024): modulation of other physiological signals, such as ECG (Charlton et al. 2017) or PPG (Chang et al. 2018, Charlton et al. 2017); chest wall volume changes and movements measured via chest straps or instrumented garments, such as Respiratory Inductive Plethysmography (Rahman et al. 2024), piezoresistive strain sensors, inertial sensors (De Fazio et al. 2021), etc.; measurement of air flow properties, such as shifts in temperature, moisture, chemical compounds (Kim et al. 2024, Yin et al. 2024), or direct measurement of respiratory airflow via a flowmeter (Massaroni et al. 2019); and lastly, acoustic sensing of respiratory sounds (Nam et al. 2015, Martin & Voix 2017, Harvill et al. 2022, Kumar et al. 2021, Romano et al. 2023, Lim et al. 2022).

Many PPG sensor-based wearable systems are commercially available, typically worn on the wrist (e.g., smartwatches), and can achieve high BR prediction accuracy when the sensor is perfectly aligned with major blood vessels in the arm. However, in practice, wrist-worn PPG sensor systems can be highly sensitive to motion artifacts

(Iorio et al. 2024, Kim & Baek 2023), leading to diminished measurement accuracy during physical activities. Similar to PPG sensor systems, which are prone to artifacts from wrist movements, ECG sensor systems, which are typically in a chest-strap form factor with electrodes built around the strap, are prone to errors from posture changes and motion (Charlton et al. 2017). Such limitations make wearable sensor systems that extract respiratory signals from the modulation of other physiological signals unsuitable for sports applications.

Chest straps or smart garments are among the most explored modes of respiratory sensing via wearable systems in the literature, and can offer high BR prediction accuracy for non-contact sports such as running and cycling (Harbour et al. 2021). However, due to the mechanism underlying this monitoring mode, motion artifacts can pose challenges to measurement accuracy, depending on the type of activity undertaken by the user (Lanata et al. 2009, Zhang et al. 2012, Vanegas et al. 2020). Beyond measurement robustness, there are also concerns about wearability and user comfort in sports settings (Zhang et al. 2025, Hu et al. 2024), motivating the exploration of alternative sensing methods better suited to sports monitoring.

Airflow-property sensing methods, including temperature, humidity, and gas-based sensing approaches, provide a more direct representation of respiration and may also enable analysis of breath chemistry. However, their practical deployment in sports settings may be constrained by their sensitivity to environmental conditions (Vitazkova et al. 2024, Liu et al. 2022).

And that leaves us with audio-based respiratory sensing systems. Some form factors suggested for audio-based respiratory sensing systems in the literature include wireless headphones (Kumar et al. 2021), “earables” (in-ear microphones) (Martin & Voix 2017, Hu et al. 2024, Zhang et al. 2025, Liu et al. 2025), and facemasks/mounted headgear (Romano et al. 2023, Lim et al. 2022, Teličko & Jakovičs 2024). In contrast,

we have embedded the acoustic sensing unit inside a mouthguard, an existing piece of protective sports equipment used in contact sports, allowing our system to be worn on-field in contact-sport settings. Peak detection has been found to be the most commonly applied method to extract BR and breathing phase from breathing audio in the literature. This chapter presents the validation of our instrumented mouthguard for breath-by-breath BR monitoring, using a peak-detection algorithm applied to the mel-spectrogram of breathing audio.

5.2 Methods

5.2.1 Data

For the study, a subset of 18 participants, 9 male participants aged 18 to 43 and 9 female participants aged 22 to 40, were taken from the dataset created in Chapter 4.

The participants performed Yo-Yo intermittent recovery level 1 (Yo-Yo IR1) tests (Krustrup et al. 2003), a running shuttle test designed to simulate workloads in sports characterized by intermittent high-intensity efforts, such as football, rugby, basketball, and similar (Bangsbo et al. 2008).

During the Yo-Yo tests, participants wore a personalised mouthguard with a built-in microphone, as described in previous studies (de Almeida e Bueno et al. 2021, Cheng & Bergmann 2023), to record breathing audio. They also wore a COSMED K5 (COSMED, Rome, Italy) portable metabolic cart, which provided the gold-standard breathing rate for system evaluation. The tests were video recorded to identify the start of each Yo-Yo test in the audio data to the nearest second. In the COSMED K5 data, the onboard GPS was used to identify the start of each test to second-level resolution, and the starting point was measured to the nearest breath. The audio and COSMED K5 data were then aligned at the start of each Yo-Yo test.

Since the COSMED K5 and instrumented mouthguard are not synchronized at the clock-level, and both data sets contain noises and artifacts, it would be considerably difficult if not impossible to compare BR measured by the two system at a breath-by-breath level. Thus a semi-random subset of the audio clips were sampled from the overall audio dataset, and was manually annotated in order to evaluate the breath-by-breath performance of the system. A random 10s long clip was sampled from every minute of the overall set. The open-source audio annotation tool - Praat (Boersma 2001) - was used to label every breathing event (inhale/exhale) in the audio subset.

5.2.2 Audio Processing

The breathing audio files were segmented into 10s long clips with 5s overlaps. The raw audio clips, which were recorded at 8kHz, were first transformed via a Short-Time Fourier Transform (STFT) algorithm with a frame length of 128, a frame step of 80, and a Fast Fourier Transform (FFT) size of 256. The STFT spectrograms were then further transformed into mel spectrograms with 20 bins and an upper-frequency cutoff at 2kHz which is the upper limit of respiratory sound frequency heard at the mouth (Sarkar et al. 2015), and a lower-frequency cutoff at 100Hz to remove stepping noises (de Almeida e Bueno et al. 2021). The mel spectrogram is then averaged across all frequency bins to produce an envelope. A 0.2s wide moving average is then applied to the envelope to remove high frequency noises while retaining information about the breathing events. Finally, a trough detection algorithm is applied to the smoothed envelope to predict respiratory flow reversal points (the start and end points of breathing events). Figure 5.1 shows an example of the algorithm at work.

5.2.3 Data Analysis

To evaluate the breath-by-breath detection accuracy, the algorithm was examined in two ways against the manual annotations. The time-lag between the predicted values

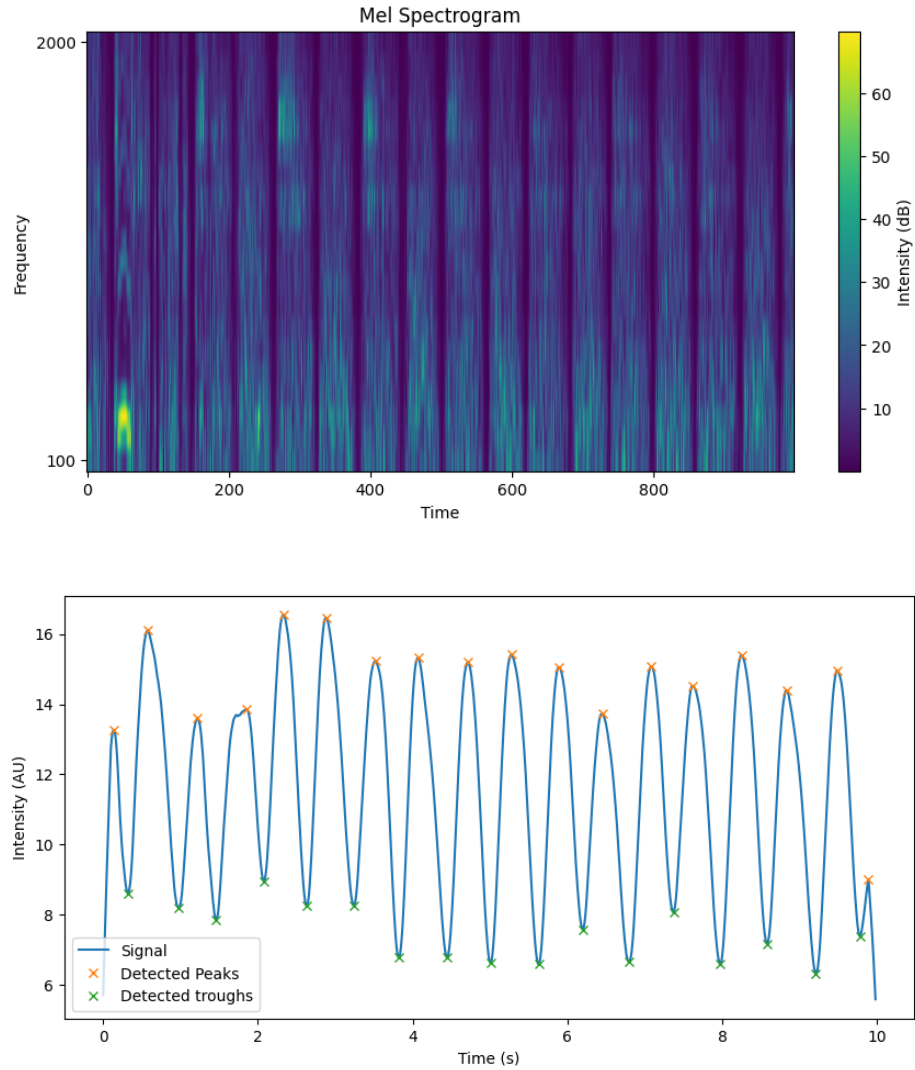


Figure 5.1: An illustration of the audio BR algorithm. The top image shows the mel spectrogram of a 10s breathing audio clip; the bottom image shows the sound envelope extracted from the spectrogram as well as the respiratory flow reversal points detected by the trough detection algorithm (the highlighted troughs on the plot).

and the labels was calculated. False positives and false negatives were counted up and precision and recall were computed.

To compare the COSMED K5 measured BR and the BR extracted from breathing

audio, both data sets were first interpolated to 10Hz, then a 10s-wide moving-average filter with a 1s step was applied to them to smooth out the curves (de Almeida e Bueno et al. 2021). A cross-correlation was then used to algorithmically align the processed audio BR, with the processed COSMED BR for each individual Yo-Yo test. Once aligned, the agreement, mean absolute error (MAE), and mean absolute percent error (MAPE) were evaluated. The agreement between the COSMED BR and the audio BR was tested through the use of the Bland-Altman plot, with the mean of differences (MoD), and the limits of agreements (LoAs) computed as a quantitative measure of agreement.

5.3 Results

In total, 230 audio clips were annotated, and in them, 3555 breathing respiratory flow reversal points were found. When evaluating the audio BR algorithm against the manually annotated labels, a total of 43 false positives and 69 false negatives were found, which equates to a precision of 98.1%, and a recall of 98.8%. The mean absolute time-lag was found to be 0.03s.

The COSMED BR has a distribution of $mean = 51.8 \pm 13.0$ and the audio BR has a distribution of $mean = 51.6 \pm 12.8$, and upon initial visual inspection, they do appear similar (see figure 5.2). However, a two-sample KS-test narrowly rejects the two distributions being statistically the same with a p-value of 0.044, and places the audio BR distribution slightly higher than the COSMED BR distribution. And indeed, when the number of breaths recorded by each system was counted, the audio BR algorithm reported 2.3% more breaths than the COSMED.

The COMSMED BR moving average was plotted against the audio BR moving average (see figure 5.3), and the mean absolute error (MAE), the mean of differences (MoD), and the limits of agreements (LoAs) were calculated. A fair agreement

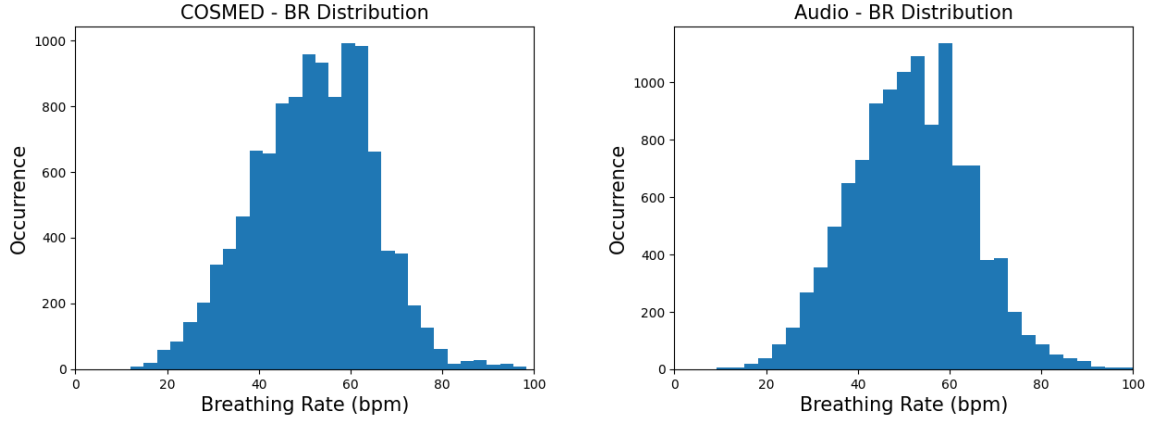


Figure 5.2: Left: distribution of BR measure by the COSMED, $mean = 51.8 \pm 13.0$; right: distribution of BR measure from the audio data, $mean = 51.6 \pm 12.8$.

between the COSMED K5 and the audio data was found, with an $MoD \pm LoA$ of -0.1 ± 9.3 , an MAE of 3.0bpm and an MAPE of 7.2%.

5.4 Discussion & Conclusion

This study is, to our knowledge, the first to estimate BR using a wearable microphone system continuously throughout high-intensity workouts under natural outdoor conditions. The findings demonstrate that respiratory audio recorded from a mouthguard-mounted microphone can be used to detect breathing events and estimate breathing rate during high-intensity outdoor exercise. Against manual annotations, the algorithm achieved high event-detection performance, with precision and recall above 98% and a mean absolute timing error of 0.03 s. When compared with COSMED K5 breathing-rate measurements, the audio-derived breathing rate showed minimal systematic bias, as demonstrated by a near-zero mean of differences (MoD).

With an $MoD \pm LoA$ of -0.1 ± 9.3 when compared to the COSMED K5, our

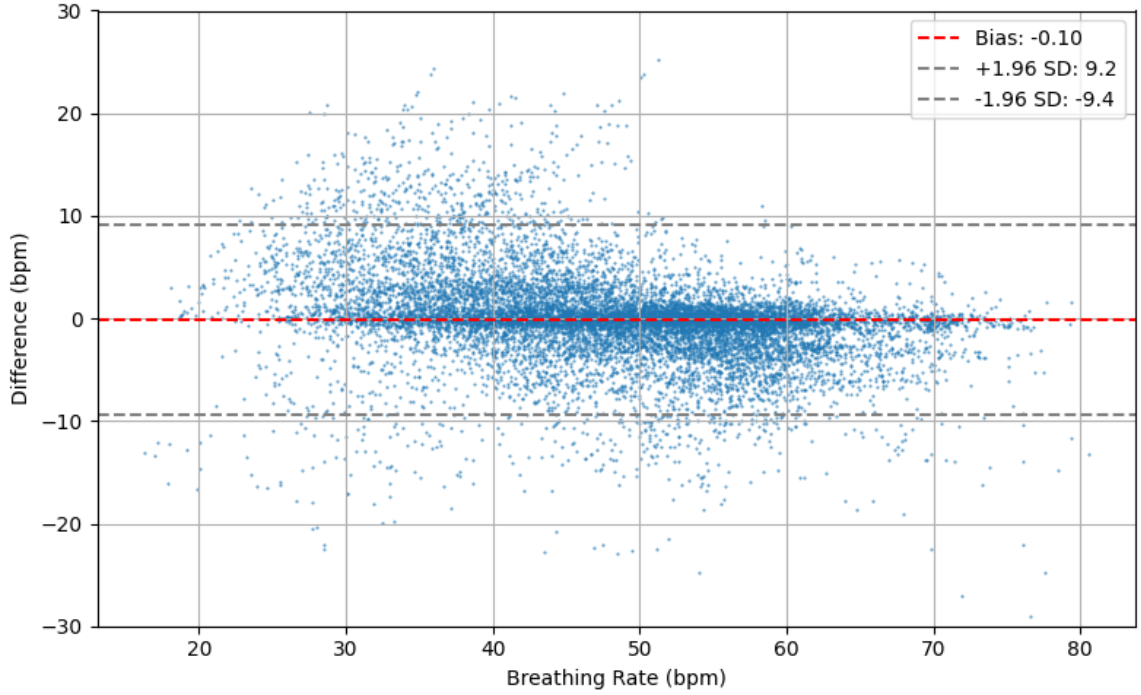


Figure 5.3: A Bland–Altman plot of the moving average of BR measured the audio data V.S. the moving average of BR measured by the COSMED. The agreement between the two appears to be fair overall with a near zero bias. Though, the audio BR algorithm appears to slightly overestimate BR at lower frequencies.

model does appear to perform comparably to a similar study carried out under indoor conditions (Nicolò, Montini, Girardi, Felici, Bazzucchi & Sacchetti 2020), where an $MoD \pm LoA$ of -2.1 ± 8.9 was reported for their 12km/h running test (which is around the speed at the start of the Yo-Yo IR1 test (Bangsbo et al. 2008)). When compared to the “earable” solution presented by Liu et al. (2025), the instrumented mouthguard recorded a higher MAE but lower MAPE due to the higher range of breathing rates in our data, which resulted from higher-intensity exercise. Our system also performed similarly to a commercially available plethysmography device, which was tested against a COSMED K4b2 in a high-intensity running test and found an $MoD \pm LoA$ of 0.2 ± 8.3 (Kim et al. 2013). Though some newer plethys-

mography devices have been shown to outperform our system (Harbour et al. 2021) under controlled conditions.

The circumstances of the Yo-Yo tests in this study are more representative of on-field monitoring conditions than lab-based tests previously presented in the literature. Appendix D shows time series plots for each individual Yo-Yo test. There, it can be seen that the COSMED and the audio data follow similar progressions over time. Higher disagreements between the audio BR algorithm and the COSMED K5 are observed for lower rates of breathing at the start of the Yo-Yo tests, which is evident in both in the Bland-Altman plot and the timeseries plots. This could potentially be due to the fact that breathing is quieter at the start of each test, as hypothesized by Nicolò, Montini, Girardi, Felici, Bazzucchi & Sacchetti (2020) as a potential source of error in the audio-based BR estimation approach under outdoor conditions. Teličko & Jakovičs (2024) reported similar findings where the measurement accuracy of their system improved during higher-intensity exercise, further validating this theory. During the manual annotation, it was noted that some breathing sounds are modulated by running strides, causing the participants to let out a double inhale or exhale. This phenomenon could also be responsible for some of the overestimation of breathing rate by the algorithm. Liu et al. (2025) presented an algorithm to decouple locomotor sounds and respiratory sounds, an approach that should be considered in future research.

Moving forward, a machine learning approach should be considered. A classification model can be employed to distinguish between inhale and exhale sounds (Harvill et al. 2022). To do this will require collecting breathing audio data perfectly synchronized with a spirometer which can be used to label the breathing phases. Perhaps a smaller dataset collected under more controlled lab conditions can be created first and used to train up an initial model, and a semi-supervised learning approach can be employed to slowly build up a larger, more diverse, and more real-world like dataset.

Chapter 6

Estimating Rating of Perceived Exertion From Cardiorespiratory Signals Using Machine Learning

Chapter Abstract

This chapter investigates whether respiratory variables add value beyond heart rate for estimating the rating of perceived exertion (RPE) from physiological signals using machine learning. The study aims to determine the relative importance of respiratory and cardiovascular features for RPE prediction and to identify an appropriate modelling approach for this task. Data were collected from 20 participants (18–43 years) during Yo-Yo Intermittent Recovery Level 1 tests while wearing a COSMED K5 portable metabolic system and a Polar H10 heart rate monitor. RPE was recorded throughout each test, and three regression models—linear regression, random forest, and multi-layer perceptron—were evaluated using eight input features: heart rate, minute ventilation, respiratory frequency, oxygen consumption, age, gender, weight, and height. Model performance was assessed using leave-one-subject-out

cross-validation, and feature importance was examined using permutation feature importance. The random forest model was found to be the most accurate, with a root mean square error of 1.849, and a mean absolute error of 1.461 ± 1.133 . Across all models, minute ventilation emerged as the most influential predictor of RPE, followed by heart rate. The chapter's key contribution to knowledge is the demonstration that respiratory information—particularly minute ventilation—is a strong predictor in machine-learning-based estimation of perceived exertion, and should therefore be considered alongside cardiovascular data in the design of future wearable monitoring systems.

6.1 Introduction

In sports, fatigue is commonly defined as “sensations of tiredness and associated decrements in muscular performance and function” and is a complex phenomenon that encompasses both physiological and psychological factors (Abbiss & Laursen 2005). The build-up of fatigue is considered a risk factor for injuries in sports, and prescribing sufficient training regimen to produce positive training outcome while minimising fatigue-related risk of illness and injury is a challenge constantly faced by coaches, sport scientists and medical personnel alike (Halsen 2014, Slobounov 2008, Jones et al. 2017, Cheng & Bergmann 2022). Various monitoring techniques have been devised by sport scientists to capture exercise workloads as ways to address these issues, which can be further divided into two categories; (i) external workloads which are measures of the physical tasks performed by athletes, often captured via local/global positioning systems, inertial sensors, and camera-based player tracking systems; and (ii) internal workloads which are the physiological and psychological responses an athlete has towards the external workloads, most prominently captured via heart rate (HR) and ratings of perceived exertion (RPE) (Halsen 2014, Cheng & Bergmann 2022, Eston 2012).

RPE is one of the most commonly applied and investigated internal workload measures in team sports, thanks to its intuitiveness and ease of use, as well as its ability to capture both physiological and psychological factors that impact exertion (Cheng & Bergmann 2022, Eston 2012). There are many variations of RPE scales developed over the years, but they all work on the same principle, consisting of a numerical scale with verbal descriptor corresponding to different levels of exertion. The two most popular RPE scales are the Borg 6-20 Category Scale and the Borg Category-Ratio-10 (CR-10) Scale (Borg 1998, Eston 2012). One major shortcoming of RPE is that it is not very suitable for continuous monitoring of training and match

workloads, since RPE is captured via active feed-back provided by the monitored athletes. Rather, RPE is often captured as a one-off measurement after a training session or match in the form of session RPE that aims to capture the overall exertion level of the session.

In terms of continuous on-field monitoring of athletes, the most commonly captured physiological measurements consist of heart rate (HR) and various HR-derived metrics, partly thanks to a high level of accessibility to commercially available validated sensor systems (Cheng & Bergmann 2022, Eston 2012). HR-based metrics are often more difficult to interpret than RPE and require greater physiological expertise to analyse (Geurkink et al. 2019). In pursuit of a continuous monitoring metric which is easy to interpret, researchers have been attempting to estimate RPE using physiological measurements that can be continuously monitored via sensor systems. (Carey et al. 2016, Vandewiele et al. 2017, Chowdhury et al. 2019, Albert et al. 2021, de Almeida e Bueno et al. 2021, Zhao et al. 2023, Rossi et al. 2019).

Much of the existing literature on estimating RPE using sensor systems leverages HR-based measures and movement data. (Carey et al. 2016, Vandewiele et al. 2017, Chowdhury et al. 2019, Albert et al. 2021, Zhao et al. 2023, Rossi et al. 2019). RPE has been theorized to be dependent on both cardiovascular and respiratory factors (Eston 2012), and several studies highlight the strong correlation between respiratory measures and RPE (Nicolò, Massaroni, Schena & Sacchetti 2020, de Almeida e Bueno et al. 2021). As such, this study aims to investigate the importance of respiratory measures in comparison to HR for estimating RPE using machine learning methods.

6.2 Materials and Methods

6.2.1 Participants

For the study, a sub-set of 20 participants with heart-rate data recorded were selected from the dataset created in chapter 4. Of the 20 participants, 10 were female and between the ages of 19 to 40, with a weight distribution ranging between 50kg and 87kg, and a height distribution ranging between 150cm and 182cm. The other 10 participants were male and between the ages of 18 to 43, with a weight distribution ranging between 63kg and 92kg, and a height distribution ranging between 170cm and 194cm.

The participants were invited to two identical Yo-Yo intermittent recovery level 1 (Yo-Yo IR1) data-capture sessions, scheduled weeks apart. Of the 20 participants, 3 had one of their runs omitted due to equipment failure on the day of data capture, and 1 was unable to attend a second session due to scheduling conflicts. Table 6.1 shows the max Yo-Yo test level and max RPE achieved by the participants in each data capture session.

6.2.2 Data Collection

In each data capture session, the participant performed a Yo-Yo intermittent recovery level 1 (Yo-Yo IR1) test (Krustrup et al. 2003) on a hard-surface outdoor multipurpose sports court. A modified version of Borg’s Category Ratio 10 scale (CR10) (Borg 1990, Foster et al. 2001, de Almeida e Bueno et al. 2021) was utilized to rate perceived exertion (RPE), see table 4.1. The RPE level of the participant was recorded at the start and finish of each shuttle of the Yo-Yo test (see Chapter 4).

Each participant wore a COSMED K5 (COSMED, Rome, Italy) portable metabolic machine, and a Polar H10 (Polar Electro Oy, Kempele, Finland) heart rate moni-

tor throughout the test, enabling the capture of respiratory frequency (Rf), minute ventilation (VE), volume of oxygen consumed (VO₂) and heart rate (HR). In this experiment the Polar H10 was linked to the COSMED K5 via Bluetooth, and the COSMED K5 was set to breath-by-breath mode for data collection.

6.2.3 Data Preprocessing

The RPE and COSMED data were recorded as time series spreadsheets and synchronized via the COSMED K5's built-in GPS which allowed for the identification of the starting point of the Yo-Yo test. In the HR data, null values were present, likely due to motion artifacts. This artifact affects around 3% of the data, all impacted data containing a null value heart rate were omitted during pre-processing. The data was then divided into 10-second clips with no overlaps, and the average Rf, VE, VO₂, HR, and RPE were obtained from these clips. The average Rf, VE, VO₂, and HR values were then grouped with the participants' physical characteristics: height, weight, gender, and age, as a set of features that were available to estimate the RPE.

6.2.4 Machine Learning Model Selection

Three regression models were tested using this set of 8 features, a linear regression (LR) model, a random forest regression model (RFR), and a multi-layer perceptron(MLP) regression model. The linear model was selected as a performance baseline as it is the simplest model possible given the nature of the data. Both the RFR model and the MLP model are non-linear models with their strengths and weaknesses. The RFR model works via generating a specified number of decision trees that form random connections between the input features. Each tree will learn a different pattern. The output is an average across all decision trees. The MLP model uses hidden layer with neurons that are connected to other hidden layers (see Chapter 1), the neurons are gated by non-linear activation functions which allow the model to learn non-linear

relationships. In theory, the MLP model is capable of learning more complex patterns than the RFR model, but would also require a larger dataset to reach high accuracy.

Both the linear model and the RFR model were built using the scikit-learn python package (Pedregosa 2011). The MLP was built using the TensorFlow library in Python 3 (Abadi et al. 2015) and consists of two dense layers with ReLU activation and a regression layer with a single fully connected node and a linear activation function. The hyperparameters of each model were tuned, and the generalizability and performance of each model were tested using leave-one-subject-out cross-validation (LOSOCV). With the optimized hyperparameters, each model was then trained on all available data to estimate feature importance.

6.2.5 Feature Importance

Permutation feature importance (PFI) is a model inspection technique used to evaluate the significance of individual features to the prediction accuracy of a model (Breiman 2001). PFI works by evaluating a trained model using randomly shuffled values for each feature, simulating scenarios where each feature, in turn, provides no information to the model, thus measuring the impact of each feature on the accuracy of the model. Unlike parameter-based importance which can only be applied to certain models (such as impurity-based feature importance for RFR), permutation feature importance is model agnostic and can be applied to any model. In this study, PFI was used to inspect the importance of Rf, VE, VO₂, HR, height, weight, gender, and age when estimating RPE using Linear, RFR, and MLP models.

6.2.6 Statistical Analysis

For model evaluation, a LOSOCV approach was taken. In LOSOCV, each cross-validation iteration involved training the model on data from all but one participant,

which is then used to validate the model (Gholamiangonabadi et al. 2020). The models were evaluated on root-mean-squared-error (RMSE) and mean-absolute-error (MAE). The results of the LOSOCV were used to compare the performance of the tuned models.

PFI was used for each model to estimate the impact of individual input features for estimating RPE. PFI was repeated 10 times for each model to minimize bias. The output of a PFI calculation is the decrease in accuracy score of the model when each feature is randomized.

6.3 Results

The maximum CR-10 RPE reached across the participants ranged from 7 to 10, with 14 participants reaching level 10, every participant's max RPE level is detailed in Table 6.1.

6.3.1 Model Validation

For the RFR model tuning, the number of estimators was tuned between 50 to 200 at an increment of 50, and 100 estimators were found to be the most optimal. For the MLP model, the number of dense layer units (32, 64, 128), the learning rate (1e-3, 1e-4, 1e-5), and the epoch number (5-20) were tuned, and the final model was built with 128 units per dense layer, trained at 1e-3 learning rate for 15 epochs.

Table 6.1 shows the detailed test results of the leave-one-subject-out cross-validation. The RFR model is the most accurate overall, with a RMSE of 1.849 and MAE of 1.461 ± 1.133 when averaging across all subjects. The RFR model was the second most consistent model across all subjects in the cross-validation, with a median of 1.528 and an interquartile range of 0.874 in RMSE, and a median of 1.261 and an interquartile range of 0.779 in MAE.

Metadata						Linear		RFR		MLP	
Participant	R1 Yo-Yo	R1 RPE	R2 Yo-Yo	R2 RPE	N	RMSE	MAE	RMSE	MAE	RMSE	MAE
D01	12.2	10	12.2	9	17	1.774	1.509	1.143	0.924	1.468	1.255
D02	14.2	10	14.3	7	50	1.595	1.266	1.137	1.004	1.618	1.142
D03	13.4	10	14.3	10	32	2.377	2.103	2.211	1.903	2.570	2.229
D04	14.2	8	15.1	9	64	1.312	1.155	1.449	1.266	1.433	1.211
D05	NA	NA	17.5	10	73	0.785	0.653	1.020	0.848	1.089	0.950
D06	17.8	10	19.4	10	177	3.403	3.260	2.047	1.692	2.245	1.929
D07	14.3	7	NA	NA	26	3.199	3.107	1.073	0.917	0.862	0.729
D08	18.1	10	17.5	10	155	2.335	1.698	1.899	1.484	1.778	1.353
D10	14.2	10	14.2	9	67	1.394	1.112	1.243	0.918	2.404	2.142
D11	NA	NA	20.1	10	108	3.109	2.736	1.839	1.601	1.603	1.213
D13	13.2	10	13.4	10	23	0.998	0.861	1.607	1.255	1.739	1.374
D14	13.2	10	13.2	10	32	3.501	3.049	3.555	2.988	3.310	2.812
D15	13.1	8	13.4	9	37	1.599	1.413	1.433	1.080	1.441	1.168
D16	15.8	9	16.7	10	111	1.370	1.204	1.948	1.519	1.353	1.155
D17	13.3	7	14.1	10	42	2.136	1.635	2.277	1.948	2.836	2.535
D18	12.3	9	13.2	9	30	1.181	0.883	1.290	0.924	1.596	1.395
D19	13.1	8	NA	NA	15	1.174	1.012	0.867	0.735	1.752	1.559
D20	16.2	10	16.2	10	106	1.750	1.476	2.086	1.713	1.898	1.642
D21	11.2	8	12.3	10	18	0.922	0.751	1.423	1.066	1.107	0.807
D24	11.2	8	12.1	8	24	0.924	0.758	2.367	2.152	1.431	1.204
All					1207	2.263	1.798	1.849	1.461	1.882	1.511

Table 6.1: Results from the LOSOCV, N stands for the number of 10-second clips from each participant. Columns “R1 Yo-Yo” and “R2 Yo-Yo” show the max Yo-Yo text level achieved by the participant in the first round (R1) and the second round (R2) of the data capture, while columns “R1 RPE” and “R2 RPE” show the max RPE level achieved by the participant in each of these data capture rounds. Each row of the table shows the results of the cross-validation iteration where the data from the stated participant was selected as the validation set, thus left out of the training set. The last row shows the average performance of each model across all subjects. The RFR model (highlighted in green) has been shown to have the best performance overall, followed by the MLP model (highlighted in yellow, and finally the linear model (highlighted in red) in last place.

The MLP model placed second in overall accuracy but was found to have performed the most consistently in cross-validation, with a RMSE of 1.882 and MAE of 1.511 ± 1.123 when averaging across all subjects; in LOSOCV, the MLP model had a median of 1.611 with an interquartile range of 0.640 in RMSE, and a median of 1.304 and an interquartile range of 0.624 in MAE.

The linear model performed the worst out of the three, with the highest RMSE (2.263) and MAE (1.798 ± 1.374) when averaging across all subjects. Furthermore,

the linear model was the most inconsistent in the cross-validation, with a median of 1.597 and an interquartile range of 1.1785 in RMSE, and a median of 1.340 and an interquartile range of 0.953 in MAE.

6.3.2 Feature Importance

Figure 6.1 shows the detailed results of the PFI calculations for each model. For all three models, VE was identified as the most important feature when estimating RPE (with median PFI scores of 3.8 for the RFR model, 3.4 for the MLP model, and 7.2 for the linear model), followed by heart rate (with median PFI scores of 3.7 for the RFR model, 1.2 for the MLP model, and 1.9 for the linear model).

6.4 Discussion

There is a growing interest in capturing perceived exertion during physical activity as a way to optimize training and potentially prevent injury due to fatigue. Recent studies already highlighted the importance of capturing respiratory measurements in sports. This study aimed to further investigate the impact of respiratory measures when estimating RPE using machine learning methods. Heart rate is one of the most commonly captured physiological metrics in sports, and is often used as a feature for modelling RPE in the current literature (Carey et al. 2016, Vandewiele et al. 2017, Chowdhury et al. 2019, Albert et al. 2021, Zhao et al. 2023, Cheng & Bergmann 2022). A previous study found that both HR and respiratory measures can be used as predictors for RPE (de Almeida e Bueno et al. 2021). As such, this study evaluated the importance of respiratory measures when modelling RPE alongside heart rate.

In this study, three regression models were evaluated. The RFR and the MLP models performed better than the linear models, which was expected due to their ability to capture more complex relationships. It was noted that during the cross-

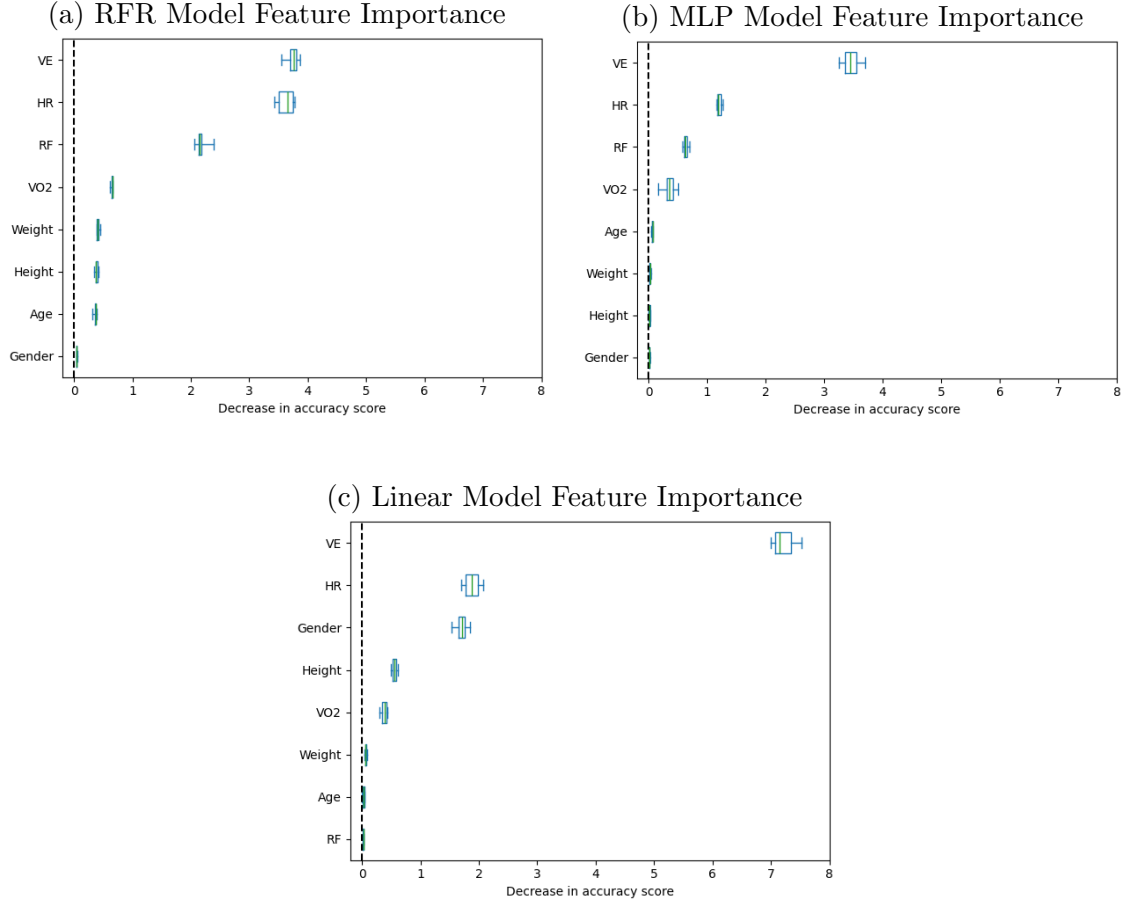


Figure 6.1: Permutation Feature Importance Boxplots for the three tested models. VE has been identified as the most important when estimating RPE, followed by HR, consistently across all three models.

validation, the MLP exhibited signs of over-fitting for several participants, which was confirmed by allowing early stopping during the cross-validation. When early stopping is enabled, the MLP's accuracy increases measurably, going from a RMSE of 1.882 and MAE of 1.511 ± 1.123 to a RMSE of 1.668 and MAE of 1.299 ± 1.046 . The early stopping method is, however, not feasible when training the model with all available data, as early stopping requires a dedicated validation set. This shows that the performance of the MLP model is limited by the cohort size of this study, and has the potential to achieve much higher accuracy with more data available. With a

larger dataset, it is also possible to further refine the hyperparameters of the MLP, which can potentially further improve the accuracy of the MLP.

The Permutation Feature Importance (PFI) calculations suggested that minute ventilation was the most impactful feature when modelling RPE across all models. Heart rate was found to be the second most important feature across these models. It was also shown that VO2 was of less importance when its margins were compared to VE and HR (VO2 PFI score median, RFR: 0.660, MLP:0.354, LR: 0.383). This finding seems to be in line with earlier findings reported by de Almeida e Bueno et al. (2021), who found minute ventilation to be better correlated ($r=0.843$) with RPE than heart rate ($r=0.770$), and both are better correlated with RPE than VO2 ($r=0.705$). These outcomes suggest that respiratory measures are crucial in understanding fatigue and warrant further investigations in sports settings. In order to capture perceived exertion, both respiratory and heart rate-based measurements should be taken into consideration.

Most studies on RPE prediction rely either on HR-based metrics or on external workload measures derived from movement data, such as GPS or IMU signals (Carey et al. 2016, Vandewiele et al. 2017, Chowdhury et al. 2019, Albert et al. 2021, Zhao et al. 2023, Rossi et al. 2019). Studies that used external workload have often been conducted on relatively homogeneous athlete populations with similar fitness levels (Carey et al. 2016, Pircoveanu & Oliveira 2024), which may limit the generalizability of their findings. Particularly, Pircoveanu & Oliveira (2024) reported that a subject-dependent approach to RPE modelling yielded better prediction accuracy than subject-independent modelling. Since different athletes would respond differently to external workload due to slightly differing fitness levels, even within a relatively homogeneous group, an individualized model calibration could reduce variability between the training and test datasets. Integrating external workload measures with cardiorespiratory measures could offer additional insights into fatigue,

and future work should consider a more holistic sensing approach with individualized model calibration.

While the collection of respiratory measurements in sports currently requires the use of expensive and obtrusive equipment, such as the COSMED K5 used in this paper, a number of recent studies have shown the possibility of collecting respiratory measures via more affordable and unobtrusive wearable systems, in the forms of chest straps (Karacocuk et al. 2019, Houssein et al. 2021), instrumented garments (Monaco & Stefanini 2021), and instrumented mouthguards (de Almeida e Bueno et al. 2021, Cheng & Bergmann 2023). These systems also provide the opportunity to directly measure or estimate VE and Rf. While measuring VO2 on-field might be impossible without obtrusive devices capable of gas analysis, it is also important to note that the results of this study suggest that VO2 might be less relevant when estimating RPE using cardiorespiratory measures. When VO2 is omitted from the models presented in this paper, they achieve a similar level of accuracy to when VO2 is included as a feature. Without VO2 inclusion, the RFR model has an average RMSE of 1.794 and MAE of 1.430 ± 1.082 (as opposed to a RMSE of 1.849 and MAE of 1.461 ± 1.133 when VO2 is also considered), the MLP model has an average RMSE of 1.956 and MAE of 1.551 ± 1.193 (as opposed to a RMSE of 1.882 and MAE of 1.511 ± 1.123), and the LR model has an average RMSE of 2.173 and MAE of 1.734 ± 1.309 (compared to a RMSE of 2.263 and MAE of 1.798 ± 1.374). These findings support the future possibility of on-field capturing of athletes' exertion levels using real-time cardiorespiratory sensing.

6.4.1 Limitations

Experimental conditions were kept as consistent as possible, whilst allowing for real-world representation. All measurements were taken outdoors at the same location. The air temperature and relative humidity were captured by the COSMED on the days of the experiments. No corrections for weather conditions were made in order

to increase the generalisability of the results. The average peak ambient temperature across all tests was $22.85 \pm 6.06^{\circ}$.

While the sensors used in this study have been shown to be valid with good accuracy in the literature, they are still prone to errors that can arise from poor contact due to motion. In the COSMED data, we detected spikes in the signal that are likely due to movements of the face mask. Furthermore, the HR data contained null values throughout the data, possibly due to temporary loss of adequate skin contact resulting from motion. The extent of the impact of these errors on the machine learning models should be further investigated in future studies as motion artifacts are inevitable in sports scenarios.

Only 80% of the volunteers completed two rounds, either due to availability conflicts or equipment failure, and around 3% of the data had to be omitted due to motion artifacts. Participants also differed in their athletic abilities, as shown by the final level reached on the Yo-Yo test. The leave-one-subject-out cross-validation was chosen, to minimize any bias due to these differences. Furthermore, the MLP model might be able to reach higher performance if more data were presented. Future studies should focus on expanding the dataset to capture more data for a wider range of participants, in order to better facilitate models that require large datasets. Our findings are based on a specific running-interval test, and their external validity may not generalize to other forms of exercise.

Lastly, the cohort size can be considered small. However, it should be noted that it was very similar in size to previous published studies that are relevant for this topic Chowdhury et al, 2019 (Chowdhury et al. 2019) (n=22); Rossi et al, 2019 (Rossi et al. 2019) (n=22); de Almeida e Bueno et al, 2019 (de Almeida e Bueno et al. 2021) (n=8); Albert et al, 2021 (Albert et al. 2021) (n=16)). The limited cohort size would likely negatively impact the generalizability of this work, for this reason, and the reasons

highlighted in the discussion above, we strongly recommend future studies to consider a larger cohort.

6.4.2 Conclusion

In this study, we investigated the feature importance when estimating RPE from cardiorespiratory measurements using machine learning methods. We compared three different regression models, linear regression, random forest regression (RFR), and multi-layer perceptron(MLP), and found that the RFR model had the best performance overall, with an average mean absolute error of 1.461 ± 1.133 in a leave-one-subject-out cross-validation. The feature importance of each machine learning model was investigated using permutation feature importance, and we found minute ventilation (VE) to be the most important feature, followed by heart rate, when estimating RPE using cardiorespiratory signals. Future works that aim to estimate RPE using wearable sensors should therefore strongly consider including respiratory data.

Chapter 7

Measuring Minute Ventilation Using an Oral Cavity Based Microphone with Deep Learning

Chapter Abstract

The previous chapter, along with studies found in the literature, highlights that minute ventilation (VE) can serve as an indicator of exertion levels in sports. Traditionally, VE is measured using spirometry, which involves bulky and intrusive equipment, making it impractical for continuous monitoring. Less obtrusive belt-style wearable sensing systems using plethysmography or ECG signals have been developed to measure VE continuously. However, these systems are susceptible to movement and tissue artifacts, limiting their effectiveness during high-intensity sports activities. Recent studies suggest that audio signals have the potential to effectively monitor breathing rates during exercise, offering high data quality. Thus, in this study, we propose a novel method for monitoring VE suitable for sports applications, utilizing respiratory audio signals captured through a microphone embedded in an oral

cavity-based device.

Data were collected from healthy volunteers during Yo-Yo Intermittent Recovery Level 1 (Yo-Yo IR1) tests using custom mouthguards equipped with embedded microphones. As a reference, ground truth measurements were obtained using a COSMED K5 portable metabolic cart (COSMED, Rome, Italy). The Yo-Yo IR1 test was selected to simulate the workloads typical of intermittent high-intensity sports. The audio recordings were aligned with the reference measurements, segmented into 8-second clips, transformed into Mel spectrograms, and used to train and compare three deep-learning regression models: a multi-layer perceptron (MLP), a recurrent neural network (RNN), and a hybrid model combining convolutional and recurrent layers.

Of the three models evaluated, the hybrid model showed the best performance in a 10-fold cross-validation, achieving a mean absolute percentage error of 9.96% and an r^2 value of 0.83. These findings indicate the potential of respiratory audio signals for effective VE monitoring in sports settings. Leave-one-out testing further showed that performance decreased for some unseen participants, highlighting the importance of wider physiological diversity in the training data. The key contribution of this work is the demonstration that respiratory audio acquired from a mouthguard can provide a viable, minimally intrusive basis for VE estimation during high-intensity exercise. This advances knowledge by introducing an alternative to torso-mounted wearable respiratory monitoring systems and establishing a proof of concept for audio-based VE monitoring in sport.

7.1 Introduction

Minute ventilation (VE), the volume of air that enters the lungs per minute, is an important respiratory parameter relevant to sports medicine. It is evidenced that VE correlates with exercise intensity (Zuurbier et al. 2009), and studies have demonstrated that VE can be a strong physiological metric for inferring an athlete’s exertion level (de Almeida e Bueno et al. 2021, Cheng et al. 2024). However, measuring VE during physical activities can be difficult. The current standard way to measure VE involves utilizing spirometry techniques with obtrusive face masks or mouthpieces (Miller et al. 2005), and in sports, this is usually captured alongside other performance metrics using a metabolic cart (Perez-Suarez et al. 2018). These obtrusive devices are only suitable for lab testing, and cannot be used in games and practices.

There have been efforts to capture VE using less obtrusive methods, both for sports purposes and for continuous patient monitoring in respiratory medicine. These methods often involve leveraging wearable sensor systems. The most popular method in the literature, by far, is to calculate respiratory airflow from chest and abdominal volume fluctuations caused by breathing (a.k.a. plethysmography) (Massone et al. 2024). This is often accomplished using belts with built-in sensors that measure the bending and stretching of the belt, such as piezoelectric strain sensors (Chu et al. 2019), capacitive sensors (Naranjo-Hernández et al. 2018), and optical fibre-grating sensors (Raicevic et al. 2016)); sensors that measure the movement of the chest and abdomen areas, such as inertial motion sensors (Karacocuk et al. 2019), and magnetometers (Houssein et al. 2021); or inductance sensors that measure the change in contact between the belt and the chest (Retory et al. 2016). This approach of monitoring respiration volume can be prone to motion and tissue artifacts, which can be particularly prevalent during physical activities (Massaroni et al. 2019, Monaco & Stefanini 2021, Massone et al. 2024, Liu et al. 2011). As is the case with these

devices, which are often designed for clinical purposes, they tend to perform best when the monitored subject is at rest, and can underperform when applied in exercise contexts (Antonelli et al. 2020). The performance of these systems can be impacted by posture, behaviours, and irregular breathing events such as coughing. All these drawbacks indicate measuring VE via plethysmography may not be well-suited for sports.

Another popular method to capture VE using wearables is to estimate VE via electrocardiogram (ECG) and other cardiovascular parameters (Cruz et al. 2020, Alam et al. 2020, Soliman et al. 2022). This method utilizes devices that are more accessible and offer better comfort to users than the former method (Massone et al. 2024), but takes a less direct approach in modelling VE using non-respiratory features. Further, ECGs are also prone to motion and tissue artifacts (Clifford et al. 2006).

Lastly, another recent development in wearable technology for monitoring respiratory volume is to design more compact spirometry devices integrated into facemasks (Zhou et al. 2020). Commercially available examples of this approach include the VO2 Master Analyzer (VO2 Master, Vernon, British Columbia, Canada), which is aimed at athletes and first responders alike. While this form factor offers a solution that can continuously measure breathing during demanding physical tasks without being affected by physical movement, it is unsuitable for on-field monitoring in contact sports.

Audio analysis for respiratory monitoring has received growing interest in recent years, especially due to its potential to monitor breathing unobtrusively. There has been a wide range of studies applying various audio processing techniques to monitor respiration frequency (Rf) under different levels of physical activity and environmental settings (Kumar et al. 2021, Harvill et al. 2022, Fahed et al. 2023, Romano et al. 2023, Cheng & Bergmann 2023, Harvill et al. 2024). Breathing audio can also be

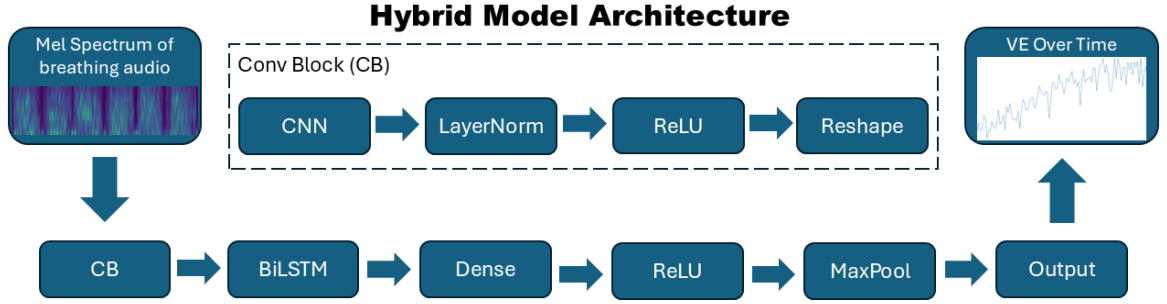


Figure 7.1: Architecture breakdown of the hybrid DNN model tested in this study. The hybrid DNN model utilizes a combination of CNN, BLSTM, and dense layers. The model takes the Mel spectrum representation of a respiratory audio clip and computes the average VE in that clip.

used to detect irregular breathing events, such as wheezing or coughing (Xue et al. 2022). Additionally, the acoustic energy of breathing audio has been used to model respiration volume in sleep and static tasks (Yadollahi et al. 2013), potentially hinting at a deeper relationship between respiratory audio features and VE. Thus, finding a method to capture VE from breathing audio under high levels of physical activity would allow for the development of a comprehensive continuous unobtrusive respiratory monitoring system that can be used out on the sport fields. To that end, we present, a proof-of-concept deep-learning-based method to measure VE during exercise using breathing audio captured via a microphone embedded inside a sports mouthguard.

7.2 Methods

7.2.1 Data

For the study, a subset of 18 participants, 9 male participants aged 18 to 43 and 9 female participants aged 22 to 40, were taken from the dataset created in Chapter 4.

The participants performed Yo-Yo intermittent recovery level 1 (Yo-Yo IR1) tests (Krustrup et al. 2003), during which, the participants were asked to wear a COSMED K5 (COSMED, Rome, Italy) portable metabolic cart, which was used to record VE in breath-by-breath (BxB) mode as the ground truth of this experiment. The participants were also asked to wear a personalized mouthguard with a built-in microphone (as described in previous studies (de Almeida e Bueno et al. 2021, Cheng & Bergmann 2023)) to record audio of their breathing during the Yo-Yo tests. The mouthguard contains an MP34DT05-A microphone (STMicroelectronics, Switzerland) set to record at 8kHz and 16 bits per sample.

The Yo-Yo tests were recorded on video for data alignment purposes, and the participants performed a synchronization task with an audio cue to synchronize the audio data with the video data. The onboard GPS on the COSMED K5 was used to identify the start of each Yo-Yo test in the COSMED K5 data. As such, the audio data is aligned with the COSMED K5 data at the start of each Yo-Yo test. The data is then sectioned into 8-second-long clips with no overlaps. The ground truth data is clipped using the timestamps recorded by the COSMED K5, which has a resolution measured in seconds and is recorded on a breath-by-breath basis at the end of each breath (an inhale-exhale cycle). The average VE of each clip is then calculated from the COSMED K5 data as follows:

$$VE_{mean} = (\sum_i T_{tot_i} \times VE_i) / T_{clip} \quad (7.1)$$

where T_{tot_i} is the duration of each breath captured in the clip, as recorded by the COSMED K5.

Due to differing physical characteristics between nose and mouth breathing, the audio data was manually annotated to separate the two. Ultimately, only mouth breathing audio was used in this study as there is a low occurrence of nose breathing in

the dataset limited to the very start of each Yo-Yo test. The resulting mouth breathing dataset has a minimum breathing rate of 19.7 bpm, meaning the 8-second-long clips contain at least 2 full breaths each. The respiratory audio clips in this study are pre-processed into Mel spectrograms for training, as it is a computationally cost-effective audio signal representation in both time and frequency domains (Purwins et al. 2019). The raw audio clips, which were recorded at 8kHz, were first transformed via a Short-Time Fourier Transform (STFT) algorithm with a frame length of 128, a frame step of 80, and a Fast Fourier Transform (FFT) size of 256. The STFT spectrograms were then further transformed into mel spectrograms with 20 bins and an upper-frequency cutoff at 2kHz which is the upper limit of respiratory sound frequency heard at the mouth (Sarkar et al. 2015), and a lower-frequency cutoff at 100Hz to remove stepping noises (de Almeida e Bueno et al. 2021). The Mel spectrograms are used as model inputs during training.

7.2.2 Deep Learning

Three deep neural network (DNN) based regression models were tested in this study. The models detailed below were built in Python (version 3.9) using the TensorFlow library (version 2.16) (Abadi et al. 2015). The KerasTuner library (version 1.4.7) (O'Malley et al. 2019) was used for hyper-parameter tuning. The hardware and firmware used for training the models was an NVidia RTX4060Ti-16GB with CUDA 11.8 and CuDNN 8.7.

Following Occam's razor while considering the nonlinear nature of this problem, a simple multi-layer perceptron (MLP) model was built as the baseline model for this study. The model consists of an input layer, hidden dense layers, a global pooling layer, and an output regression layer, where each layer is fully connected (i.e., all neurons in a layer are connected to every neuron in the subsequent layer). The hyper-parameters that were tuned for this model were: the number of hidden dense

layers (one layer, two layers); the number of neurons in each dense layer (16, 32, 64, 128); and the activation function (ReLU, eLU).

A recurrent neural network (RNN), which is a type of DNN that is purposefully designed to handle sequential data (Elman 1990), was also considered for processing the data. RNNs compute information for each time step using information at that time step as well as a hidden state computed from the previous time step, making it more capable of solving complex temporal problems than MLPs. However, RNNs are prone to vanishing gradient issues during backpropagation, making them impractical for long data sequences. To mitigate the vanishing gradient problem, memory cells and gating mechanisms are introduced to maintain a more stable flow of gradients during backpropagation, forming a special RNN architecture named Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber 1997). Bidirectional LSTMs (BLSTMs) are a powerful tool in audio processing tasks (Purwins et al. 2019). In this study, a BLSTM model, consisting of an input layer, hidden BLSTM layers, a global pooling layer, and an output regression layer, was tested against the MLP baseline model. The hyper-parameters that were tuned for the BLSTM model were: the number of hidden BLSTM layers (one layer, two layers); and the number of RNN units in each BLSTM layer (8, 16, 32, 64).

The hybrid model proposed on this study contains a block with a convolutional neural network (CNN). A CNN is a type of DNN that utilizes convolution filters (referred to as kernels in the context of a CNN model), when these filters are applied to a data array, they transform the data into “feature maps” (LeCun et al. 2015). CNNs are commonly applied in image-processing tasks such as computer vision. In audio processing applications, CNNs have been combined with RNNs to construct sophisticated hybrid models that can leverage both CNNs’ ability to extract hidden features from spectrograms as well as RNNs’ ability to learn complex temporal patterns, with the DeepSpeech2 model being a shining example of this (Amodei et al.

2016). In this study, a CNN-RNN hybrid model inspired by the DeepSpeech2 model was tested. The hybrid model consists of three core types of hidden layers, a convolutional layer, a BLSTM layer, and a dense layer, figure 7.1 shows a detailed breakdown of the construction of the hybrid DNN model. The hyper-parameters that were tuned for the hybrid model were: the number of CNN kernels (4, 8, 16, 32); the number of RNN units (8, 16, 32, 64); and the number of neurons in the dense layer (16, 32, 64, 128). The CNN kernel size was set to 10 by 3, a shape that is approximately scaled down from the kernels of the DeepSpeech2 model based on the difference in the input shapes. The CNN kernel size was not further tuned due to computational constraints.

7.2.3 Model Evaluation

All three models were evaluated in two different data configurations: i) in the first data configuration, the models are exposed to all subjects during training, and were evaluated in a 10-fold cross-validation with identical data splits across all models; ii) in the second data configuration, 7 subjects with the most data points were each individually withheld from training in a leave-one-out cross-validation. The first data configuration allowed us to analyze the generalizability of the models on athletes the device has been calibrated to, while the second data configuration gave us information on the generalizability of the models on random athletes "out of the box". Both types cross-validation helped circumvent the limited size of the dataset while providing invaluable insights into the performance of the models (Umberto 2018).

7.3 Results

In total, 1511 eight-second-long respiratory audio clips were collected from 18 participants during Yo-Yo IR1 tests. The average VE values of the data clips range

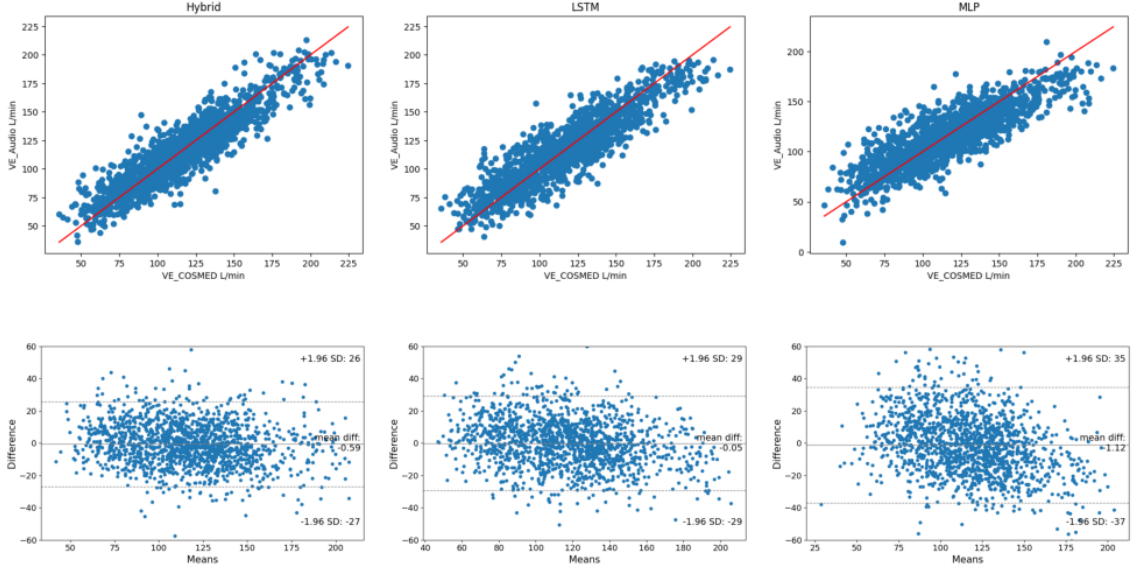


Figure 7.2: The top row shows the scatter plot of validation data pooled together across the cross-validation for each model; the bottom row shows the Bland–Altman plots corresponding to the top row.

from 35.7 L/min to 224.5 L/min, with a mean and standard deviation of 116.8 ± 32.9 L/min.

The final configurations for the tuned models were as follows. The multilayer perceptron (MLP) model was set to one hidden dense layer with 64 neurons and ReLU activation. The BLSTM model was set to one hidden BLSTM layer with 64 RNN units. Finally, the hybrid model was configured with 8 CNN kernels in the CNN layer, 64 RNN units in the BLSTM layer, and 128 neurons in the dense layer.

In the 10 fold cross-validation, root mean squared errors (RMSE) and mean absolute percentage errors (MAPE) were averaged across the cross-validation to compare the accuracy of the tested models. Coefficient of determination (r^2) values, as well as Bland–Altman plots, were used to compare the goodness of fit and agreement of the models with the gold-standard COSMED K5 data. The MLP model served as a performance baseline, with the lowest accuracy and goodness of fit; as expected,

Model	RMSE (L/min)	MAPE (%)	r^2
Hybrid	13.50	9.96	0.83
BLSTM	14.89	11.10	0.79
MLP	18.35	13.73	0.69

Table 7.1: This table shows validation errors averaged across all folds of the cross-validation, and the r^2 values for the models measured against the COSMED K5.

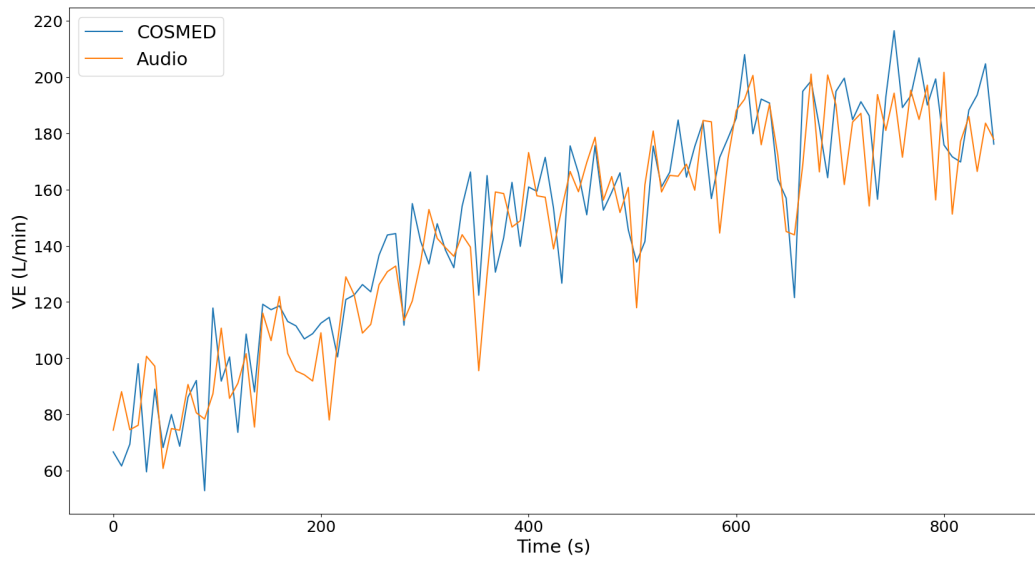


Figure 7.3: An example of applying the hybrid model to audio data collected during a full Yo-Yo IR1 test run, the estimated VE is plotted alongside the COSMED K5 data against time.

the BLSTM was better suited for time series data, performing noticeably better than the MLP; lastly, the hybrid was able to leverage all the added complexities of the differing types of hidden layers, outperforming both the MLP and the BLSTM. Table 7.1 shows the RMES, MAPE, and (r^2) values for the three models. Figure 7.2 shows the validation results pooled together from all cross-validation folds plotted against the ground truth values, as well as the Bland–Altman plots for all three models in descending performance. Lastly, figure 7.3 shows an example of when the hybrid VE model is applied to audio from a full Yo-Yo IR1 test.

For the leave-one-out cross-validation, MAPE, and RMSE were computed for each held-out subject during testing, and are compared against their subject-exposed counter parts obtained during the 10 fold cross-validation. The outcomes of this analysis are shown in table 7.2. Additionally the results of this test are shown as scatter plots for each subject in Appendix E.

Subject	Model	RMSE (L/min)		MAPE (%)	
		SE	SH	SE	SH
D05	Hybrid	13.9	17.9	8.5	10.5
	BiLSTM	14.9	20.2	8.7	12.6
	MLP	18.3	21.0	11.7	13.2
D06	Hybrid	10.7	12.7	7.7	8.7
	BiLSTM	12.8	14.3	9.0	10.1
	MLP	14.3	16.2	10.1	10.8
D08	Hybrid	11.5	14.1	8.7	10.8
	BiLSTM	13.9	14.6	10.9	11.4
	MLP	14.9	16.4	11.5	12.8
D09	Hybrid	11.8	12.8	8.3	8.9
	BiLSTM	11.7	10.2	7.4	6.8
	MLP	13.9	19.2	8.7	12.5
D11	Hybrid	16.8	33.3	10.2	17.6
	BiLSTM	17.6	36.1	11.0	19.2
	MLP	22.9	35.6	13.5	19.7
D16	Hybrid	10.7	10.0	7.0	7.1
	BiLSTM	11.1	10.9	7.5	7.6
	MLP	10.8	11.1	7.7	7.9
D20	Hybrid	14.2	23.8	11.1	19.5
	BiLSTM	17.5	24.5	13.8	19.6
	MLP	25.0	32.8	20.6	26.0

Table 7.2: Results from the leave-one-out cross-validation, where SE and SH stand for the two cross-validation data configurations. SE stands for subject exposed and denotes the 10-fold cross-validation models; SH stands for subject held-out and denotes the leave-one-out cross-validation models.

7.4 Discussion

7.4.1 General Discussion

In this study, we presented a proof-of-concept VE monitoring method using breathing audio and deep learning. Out of all the models tested, the overall best-performing was a hybrid deep-learning model that utilizes CNNs’ ability to extract features from 2D images like the audio spectrograms used as inputs for the models, and RNN’s ability to learn from complex time-series data.

While the hybrid deep learning model has performed consistently well across different subjects in the 10-fold cross validation test where the models are exposed to all subjects during training, the model performed poorly (with greater errors) on two subjects (D11 and D20) during the leave-one-out cross-validation test. Further analysis revealed that subject D11 exhibited a wider range of VE values than other participants, including a notably higher maximum VE. The large errors for D11 occurred primarily when their VE exceeded the range observed in the training data (see figure E.5), suggesting that the model functions well within the training distribution but struggles to extrapolate beyond it. In fact, when analyzing only the periods where D11’s VE remained within the typical range, the errors (MAPE = 13.5%, RMSE = 18.7) were comparable to those of other participants. In contrast, the elevated errors for subject D20 are less easily explained. An interview revealed that D20 is a long-distance runner who uses a specialized breathing technique, distinguishing their breathing patterns from those of the other subjects. This, in contrast to the rest of the subjects, underscores the importance of incorporating a diverse range of physiological profiles in the training dataset to ensure better generalization.

In a previous study, we showed that high-quality breathing data can be collected via a mouthguard integrated with a microphone (Cheng & Bergmann 2023). Building on that work, this study demonstrates the potential of using breathing audio to

measure VE even during high-intensity exercise. In addition to benefitting from the overall data quality, integrating a sensor system into a mouthguard also improves the chance of player compliance, as they are often required to wear a mouthguard for protective reasons.

Further, the method of capturing VE via breathing audio can have applications outside of sports. VE along with its components, respiratory frequency (Rf) and tidal volume (VT), are of great interest in respiratory medicine (Massaroni et al. 2019, Monaco & Stefanini 2021, Massone et al. 2024). With appropriate modifications to the form factor of the wearable system (such as mounting the microphone on the outside of the body near the respiratory tract to further reduce the obtrusiveness and increase user comfort), respiratory audio can potentially be used for continuous patient monitoring in a telemedicine setting. And rather than as a standalone system to measure VE, breathing audio features can be used to supplement existing methods of VE monitoring via wearable, which may lead to improvements in detection accuracy.

7.4.2 Error Analysis

There are several potential sources of error in labelling the audio data, the first of which is the fact that the audio data and the COSMED K5 data were not synchronized at a hardware level. With the COSMED K5 recording data on a per-breath-basis, the audio data and the COSMED K5 data could be out of sync by up to a full-breath. This dataset has a minimum Rf value of 19.7 bpm, which means the data and the label can be misaligned by up to 3 seconds. The error can be somewhat mitigated using the alignment adjustment made using cross-correlation between breathing rates measured from the audio data and the COSMED as detailed in chapter 5, though there is no guarantee that method had indeed yielded the true alignment between the COSMED and the audio data.

A second source of error comes from the way the COSMED K5 captures a breathing cycle. Typically, a breathing cycle can be segmented into inspiration, post-inspiration pause, expiration, and post-expiration pause. The COSMED K5, however, marks the transition between inspiration and expiration based on the change in the spin direction of the built-in turbine, which often fails to fully capture respiratory pauses due to angular momentum in the turbine. Thus, if there is a respiratory pause at the beginning or the end of an audio clip, the calculated average VE label would not be able to account for that. The respiratory pause duration is correlated with the breathing cycle duration but can be subject to the physical task performed during the breathing cycle (Boiten 1993), as such, it is difficult to estimate the level of uncertainty in the labels caused by this issue.

And lastly, the COSMED K5 can experience unexpected motion artifacts. Such artifacts can cause occasional rapid spikes in the data. This type of error is mostly mitigated when averaging the VE value over time, but can still have a small negative impact on the label accuracy in extreme cases.

7.4.3 Limitations and Future Studies

As mentioned at the beginning of the error analysis section, the clocks of the COSMED K5 and the acoustic sensor system are not perfectly aligned, leading to uncertainty in the label accuracy. Future investigations should attempt to build a test rig with fully synchronized clocks at a hardware level. Fully synchronized systems will also allow for breath-by-breath labelling of the audio data, which would in turn address the second issue raised in the error analysis. Further, this would allow for the development of breathing phase detection and VT detection algorithms on a per-breath-basis.

Only mouth breathing audio data was used in this study due to limited amount of nose breathing data being present in the set. However, it was noted that nose

breathing data behaves differently from mouth breathing data. A larger set of nose breathing audio data should be captured in future studies. A classifier can be constructed to distinguish between nose and mouth breathing data, and separate VE detection models can be trained for the different modes of breathing.

Additionally, the method presented in this study has only been validated for running exercise, further investigations are needed for other physical activities. Moreover, it was noted by the author during data pre-processing that running and walking show different noise profiles. A similar study that presented a method to monitor VE using ECGs found that the detection accuracy can be increased if the context of the activity is provided to the model (Alam et al. 2020). This finding could potentially apply to audio as well, meaning future studies should investigate integrating IMU data for activity classification into the audio-based VE algorithm presented in this study.

7.5 Conclusion

This study presents a novel method of measuring VE from audio signals via deep learning. Respiratory audio signals can be collected unobtrusively via microphones embedded inside custom sports mouthguards. Incorporating the sensor system into an existing piece of sports equipment, as such, allows this method to be translated to on-field sports monitoring in the future with a good likelihood of compliance from the players.

This study serves as a proof-of-concept for monitoring VE with audio. While this concept has been demonstrated for sports monitoring here, with appropriate adjustments to the wearable device form factor and a more comprehensive dataset, this technology has the potential to be applied to broader fields of medicine.

The hybrid model presented in this study demonstrated great potential for extracting respiratory information from audio, utilizing CNNs' ability to extract fea-

tures from spectrograms and BLSTMs' ability to detect patterns in the time domain. Further studies should attempt to fully leverage BLSTMs' time domain computation capabilities and aim to measure respiratory parameters on a breath-by-breath basis.

Overall, the method presented in this study differs greatly from the torso-mounted wearable systems popular in the current literature (Massone et al. 2024, Monaco & Stefanini 2021), prompting a new approach to measuring VE using wearables.

Chapter 8

Discussions & Conclusion

8.1 Thesis Summary & Retrospective Thoughts

This project set out to develop a method of on-field player monitoring for team contact sports - one that can provide new insight into player performance and exertion level in real-time, enable new research in sport physiology, and hopefully lead to player training and management changes that can improve the health and well-being of athletes in the future.

To achieve this aim, we began this project with a review of current methods and recent advancements in on-field sports monitoring in Chapter 2. This review revealed that most monitoring practices fall into two main categories: workload monitoring and impact monitoring (Cheng & Bergmann 2022). Research in impact monitoring has primarily focused on detecting head injuries by measuring head acceleration using wearable sensor systems (O'Connor et al. 2017). Despite extensive efforts, these devices have not greatly improved athlete health outcomes, largely due to their high error rates and limited accuracy in diagnosing concussions (O'Connor et al. 2017, Cheng & Bergmann 2022, Haste et al. 2025). Many consider these devices to be fundamentally flawed, as they measure acceleration of the skull rather than the brain,

and so are unable to detect brain trauma.

Workload monitoring captures both the physical demands placed on athletes (external workloads) and their physiological and psychological responses (internal workloads) (Bourdon et al. 2017). External workloads are typically measured using technologies such as GPS, LPS, cameras, or IMUs to track player movement during games (Benson et al. 2020). This data is valuable for designing training programs that better replicate game conditions (Oliva-Lozano et al. 2021). Internal workloads, while more complex to assess, include physiological indicators like heart rate, oxygen uptake, and core temperature, as well as subjective measures such as rating of perceived exertion (RPE) and wellness questionnaires. Among these, RPE and heart rate are the most commonly used outside of lab settings (Cheng & Bergmann 2022). Research suggests that perceived exertion correlates with player wellness (Perri et al. 2021, Clemente et al. 2018), and when combined with external workload data, may help forecast injury risk (Vallance et al. 2020). However, due to RPE being a survey-based metric, it is impossible to measure RPE continuously on-field.

Respiratory measurements have long been used to measure the fitness level of athletes in lab settings (Löllgen & Leyk 2018), but have seen little use in internal workload monitoring (Nicolò et al. 2017). Two respiratory measurements, BR and minute ventilation (VE), have been shown to correlate closely with perceived exertion (de Almeida e Bueno et al. 2021, Nicolò, Montini, Girardi, Felici, Bazzucchi & Sacchetti 2020). Thus, there seems to be a current gap in on-field sports monitoring — specifically, a lack of respiratory monitoring was revealed. Addressing this gap could enable continuous estimation of player exertion in real-time, directly on the field.

Since the beginning of this project, there appears to be an increasing interest in measuring BR via wearable sensors, both for sports and medical purposes. Multiple types of wearable sensor systems for capturing breathing rate (BR) have been identi-

fied in the literature. These systems generally fall into four main categories (Massaroni et al. 2019, Liu et al. 2019, Costanzo et al. 2020, Vanegas et al. 2020, Vitazkova et al. 2024): modulation of other physiological signals, such as ECG (Charlton et al. 2017) or PPG (Chang et al. 2018, Charlton et al. 2017); chest wall volume changes and movements measured via chest straps or instrumented garments, such as Respiratory Inductive Plethysmography (Rahman et al. 2024), piezoresistive strain sensors, inertial sensors (De Fazio et al. 2021), etc.; measurement of air flow properties, such as shifts in temperature, moisture, chemical compounds (Kim et al. 2024, Yin et al. 2024), or direct measurement of respiratory airflow via a flowmeter (Massaroni et al. 2019); and lastly, acoustic sensing of respiratory sounds (Nam et al. 2015, Martin & Voix 2017, Harvill et al. 2022, Kumar et al. 2021, Romano et al. 2023, Lim et al. 2022).

PPG-based systems, typically worn on the wrist like smartwatches, can achieve high BR prediction accuracy when well-aligned with major blood vessels. However, real-world use—particularly during movement—often results in poor data quality due to motion artifacts (Iorio et al. 2024, Kim & Baek 2023). Similarly, ECG-based systems, commonly implemented as chest straps with embedded electrodes, are susceptible to errors caused by changes in posture (Charlton et al. 2017). Such limitations make wearable sensor systems that extract respiratory signals from the modulation of other physiological signals unsuitable for sports applications. Chest straps or smart garments are among the most explored modes of respiratory sensing via wearable systems in the literature, and can provide accurate BR measurements in non-contact sports like running and cycling (Harbour et al. 2021). However, their form factors are generally unsuitable for contact sports, where physical collisions can shift or displace the sensors. Currently, systems that monitor breathing-induced changes in temperature and humidity are best suited to controlled lab environments, given their form factors and susceptibility to environmental factors (Vitazkova et al. 2024, Liu et al.

2022). Turbine systems and gas sensors are less prone to motion and environmental artifacts and offer high measurement accuracy, but require bulky, intrusive equipment, including tightly strapped silicone face masks, to function (Massaroni et al. 2019), making them less ideal for real-time on-field monitoring.

This leaves audio-based systems as a promising alternative. Prior work has proposed audio sensors integrated into earbuds (Martin & Voix 2017, Kumar et al. 2021) or facemasks (Romano et al. 2023, Lim et al. 2022). In our study, we proposed integrating a microphone-based respiratory sensing system into an existing piece of sports protective gear to enable unobtrusive on-field monitoring in contact sports. Chapter 3 described the pilot studies carried out to determine the device form factor, as well as the development work that followed. Two suitable sensor placement locations were identified: either mounted to the head on a helmet/pair of protective goggles, or encapsulated inside a mouthguard. We compared the signal-to-noise ratio (SNR) and respiratory rate estimation potential of these two placements. Four participants were tested in both indoor and outdoor environments, with respiratory rate estimated using a fast-Fourier transform (FFT) method. Through this study, we were able to show that the sensor encapsulated inside a mouthguard was able to collect high-quality breathing data, even under noisy outdoor conditions, thus making it potentially suited for on-field sport monitoring. We also found that the breathing audio volume is strongly correlated to air intake volume, hinting at the potential to estimate VE using breathing audio. These preliminary findings suggest that a respiration monitoring system embedded in a mouthguard holds significant promise for on-field use.

To further develop and validate a respiration monitoring smart mouthguard, a substantial amount of data is needed. Due to the unique and novel nature of our project, there wasn't an openly available dataset suited for our purpose. Thus, we designed a data collection protocol to generate a comprehensive dataset of breathing

audio captured during high-intensity exercise, providing the foundation for our continued research. Chapter 4 shows the details of this data collection. In this study, 21 healthy participants with diverse ages, sexes, and fitness levels were recruited to capture a broad range of physiological variability. Each participant completed the Yo-Yo Intermittent Recovery Level 1 (Yo-Yo IR1) test (Krustrup et al. 2003), a shuttle run specifically designed to simulate the intermittent high-intensity workloads characteristic of sports such as football, rugby, and basketball (Bangsbo et al. 2008). The tests were conducted on an outdoor sports court, where participants wore an instrumented mouthguard (see Appendix ??) that continuously recorded breathing audio. To capture ground truth data, a COSMED K5 portable metabolic cart was used to measure key respiratory metrics such as breathing rate (BR) and minute ventilation (VE), along with other physiological and workload indicators, including heart rate (via a Polar H10) and running speed (via the K5’s onboard GPS). Participants also rated their perceived exertion using the Borg CR10 scale (Borg 1990) at the start of each shuttle. After removing incomplete recordings due to occasional device failures, a final dataset from 18 participants was retained. This dataset is notably larger and more comprehensive than comparable studies in the literature, encompassing a wide range of physiological and kinematic variables.

Using this dataset, we developed algorithms to estimate BR and VE directly from breathing audio and further demonstrated the value of respiratory metrics in exertion monitoring. Though some shortcomings of the dataset were also identified in those further studies, these shortcomings should be addressed in future data collection protocols. One major flaw of this experimental design is that the COSMED K5 cannot be synchronized with the smart mouthguard on a hardware level. Steps were taken to ensure the alignment between the audio data and the COSMED K5 data was as close as possible, but the uncertainty is limited by the resolution of the COSMED K5. The COSMED K5 records respiratory data on a breath-by-breath basis, with

each measurement timestamped to the nearest second. The breaths are determined by a turbine in the facemask which can experience artifacts during running, occasionally producing false spikes in breathing rate, as seen prominently in Subject 8 Test 2 (Figure D.10). As a result, the dataset was not ideal for developing breath-by-breath VE (and consequently tidal volume, VT) detection algorithms. To minimize errors caused by temporal misalignment, I instead focused on estimating average VE across audio segments containing multiple breaths (more on this later). I was, however, able to develop and evaluate a breath-by-breath BR detection algorithm with manually annotated breathing audio data, as showcased in Chapter 5.

In Chapter 5, we demonstrated that our smart mouthguard system is capable of estimating breathing rate during high-intensity outdoor exercise. Using the dataset detailed in Chapter 4, breathing audio was transformed into mel spectrograms, converted into a respiratory sound envelope, and analyzed using a trough-detection algorithm. The algorithm is evaluated against manual annotations and COSMED K5 breathing-rate measurements. The results show that mouthguard-recorded audio can support accurate breathing-event detection and breathing rate estimation, achieving performance comparable to other studies conducted under more controlled, indoor laboratory conditions. These include systems using microphones integrated into masks (Nicolò, Montini, Girardi, Felici, Bazzucchi & Sacchetti 2020) and commercially available chest-belt devices (Kim et al. 2013).

Building on the promising results of our smart mouthguard system for BR monitoring, we extended our investigation in Chapter 6 to explore exertion modelling using cardiorespiratory measures. This work aimed to address two key objectives: (i) to determine whether RPE can be estimated from objective physiological data, and (ii) to assess the relative importance of various cardiorespiratory metrics in accurately modelling physical exertion. Our results showed that a non-linear model using a random forest regression (RFR) architecture could predict RPE with acceptable

accuracy. More importantly, across multiple models, both minute ventilation (VE) and heart rate (HR) consistently emerged as key predictors of exertion. These findings not only motivated a deeper investigation into estimating VE from audio signals, but also highlighted the importance of capturing multiple types of internal workload measures in order to comprehensively assess the multifaceted nature of exertion.

Finally, to round out this thesis, we set out to develop an algorithm that can estimate VE from breathing audio, an important step toward on-field exertion monitoring as outlined by Chapter 6. As with BR monitoring, plethysmography emerged as the most common method for measuring VE and tidal volume (VT) using wearable sensors in the literature (Massone et al. 2024). However, most existing studies employing plethysmography have focused on patients at rest or engaged in low-intensity activities for clinical applications. A major concern with this technique in dynamic settings is its susceptibility to motion and tissue artifacts, which become especially problematic during vigorous physical activity (Massaroni et al. 2019, Monaco & Stefanini 2021, Massone et al. 2024, Liu et al. 2011). Additionally, factors such as posture changes, user behaviors, and irregular breathing events (e.g., coughing) can negatively affect measurement accuracy. These limitations suggest that plethysmography may not be ideally suited for use in athletic contexts, reinforcing the need for alternative approaches such as audio-based respiration monitoring.

Using the dataset established in Chapter 4, Chapter 7 investigates algorithms for estimating minute ventilation (VE) from breathing audio. Deep learning proved to be instrumental in extracting meaningful information about VE from acoustic signals. Ultimately, a deep learning architecture was developed, comprising convolutional neural network (CNN) layers for spectrogram-based feature extraction, bidirectional long short-term memory (Bi-LSTM) layers to capture temporal dynamics, and fully connected layers to map these representations to VE estimates. In subject-exposed testing, the model demonstrated consistent performance across the

participant pool. However, leave-one-out cross-validation revealed a notable limitation: the model struggled to generalize to subjects whose breathing behaviour differed considerably from that represented in the training data. This limitation was particularly evident with Subject D11, whose peak performance exceeded that of all other participants—the model struggled to extrapolate VE values beyond the range encountered during training; and with Subject D20, a long-distance runner who employed a unique breathing technique, which confused the model throughout the tests. This highlights the need for greater subject diversity and data quantity in order to improve the generalizability of the model.

During data pre-processing, it became apparent that nose breathing exhibited distinct acoustic characteristics compared to mouth breathing, necessitating separate treatment in the modelling process. However, due to the limited amount of nose breathing data available in the dataset, the final model was developed exclusively for mouth breathing. In future work, a more extensive dataset of nose breathing audio should be collected, particularly at lower activity levels where nose breathing is more prevalent. This would enable the development of a dedicated VE estimation model for nose breathing. Additionally, a classifier could be implemented to automatically distinguish between mouth and nose breathing, allowing the system to apply the appropriate model for each case and improve the overall estimation accuracy.

Ultimately, this final chapter serves as a good proof-of-concept for VE monitoring using audio data, showcasing a novel approach that could be applied to sports settings. Future studies should build on this work using bigger and higher-quality datasets. A major limitation of this study was the fact that the COSMED K5 and the microphone board were not synchronized on a hardware level, which introduced much uncertainty to the ground truth data used to train the model. While the goal of this study was to showcase the potential to monitor respiration under harsh outdoor sport conditions, future studies should also look to collect data under a more controlled lab environment

that can be used to create a model with a higher detection accuracy, and have the model subsequently validated against the COSMED K5 under outdoor conditions. Additionally, creating a dataset where the ground truth is fully synchronized with the audio data would support breath-by-breath analysis of metrics like tidal volume (VT) and enable the development of models capable of identifying individual breathing phases, providing deeper insights into respiratory patterns and improved breathing rate detection accuracy.

8.2 Future Work

In addition to addressing the issues and limitations identified in these previous studies, there are several other important directions to explore in the future development of this project.

8.2.1 Integrations With Additional Sensors

In Chapter 6, I found that both VE and HR contribute significantly to exertion modelling. HR is currently monitored using belt devices with electrodes, such as the Polar H10 used in this project. However, the reasoning behind why sensor belts and smart garments are unsuitable for monitoring breathing in contact sports can be made for HR monitoring belts. Therefore, to truly create a wearable system capable of monitoring exertion, incorporating HR sensing into the device is crucial. de Almeida e Bueno et al. (2023) showcased a potential method to monitor HR from the oral-cavity using a PPG sensor. In addition to providing HR information, PPG sensor can also be used to augment BR detection (Charlton et al. 2017), and provide additional physiological measurements like oxygen saturation (SpO2).

A similar study demonstrated that incorporating contextual information about physical activity can enhance the accuracy of VE estimation from ECG signals (Alam

et al. 2020). This finding suggests that integrating an inertial measurement unit (IMU) into our system could be beneficial — to both classify the type of exercise being performed and provide contextual information that may improve the performance of audio-based respiratory monitoring algorithms. Furthermore, activity classification introduces the potential to estimate additional metrics, such as energy expenditure, thereby broadening the scope and utility of the system for comprehensive on-field athlete monitoring.

8.2.2 Applications in Sports

This project has demonstrated how oral-cavity-based acoustic sensing could be used to extend sports science and training practice by providing an additional internal workload measure alongside existing tools such as position-based monitoring, heart rate, and RPE. In applied settings, a microphone-instrumented mouthguard could allow practitioners to monitor breathing rate and estimate minute ventilation during high-intensity intermittent exercise, helping coaches interpret how individual athletes respond physiologically to specific drills, conditioning sessions, or match-related physical demands. Such information could support training prescription, fatigue monitoring, recovery assessment, and return-to-play decisions by identifying when an athlete’s respiratory response is disproportionate to their external workload. When integrated alongside heart rate and external load measures, high-accuracy exertion prediction is possible, allowing for more responsive decision-making during training and match scenarios that could lead to better player health outcomes.

However, the present findings should be viewed as proof of concept rather than evidence of immediate readiness for widescale deployment. Further research is needed using larger and more diverse athlete cohorts, sport-specific on-field protocols, and data collected during real training and competition, where movement artifacts, contact, speech, environmental noise, and changes between nose and mouth breathing

may affect accuracy. Algorithms and additional sensor integration solutions need to be developed to minimize possible sources of measurement errors. Wider uptake would therefore require further validation and improved model generalizability before the technology could be used with confidence for routine on-field decision-making.

8.2.3 Future Applications In Respiratory Medicine

Respiratory conditions are widespread and pose real threats to both lifespan and health span. In the UK, approximately 8% of the population are diagnosed with asthma, according to Asthma and Lung UK. However, studies have shown that asthma is frequently under-diagnosed and under-treated, particularly in older adults (Gibson et al. 2010). Furthermore, at least one-third of individuals with asthma also experience breathing pattern disorder (BPD)—also known as dysfunctional breathing—which can complicate asthma management, as conventional medication often fails to fully address the associated symptoms (Boulding et al. 2016).

Chronic obstructive pulmonary disease (COPD) is another common respiratory condition, affecting an estimated 1.2 million people in the UK, according to the NHS. COPD exacerbations are the second leading cause of emergency hospital admissions in the country, accounting for one in eight hospital admissions. Misdiagnosis of COPD as asthma is not uncommon, leading to suboptimal treatment and symptom management (Ragaišienė et al. 2019). The diagnosis of asthma and COPD remains particularly challenging due to its heterogeneous nature (Daines et al. 2021), and the diagnosis and management of COPD exacerbations often rely heavily on clinician judgment.

Digital healthcare technologies offer promising opportunities to enhance patient care for individuals with respiratory conditions. These tools can play a crucial role in detecting early signs of disease exacerbation, thereby helping to reduce health risks

and improve outcomes (Daines et al. 2021). For instance, respiratory rate monitoring has demonstrated the ability to predict severe nighttime asthma attacks during sleep (Nagakumar et al. 2024). Likewise, growing evidence supports the use of respiratory rate and volume data to identify COPD exacerbations (Yañez et al. 2012, Mohktar et al. 2015). These studies underscore the value of integrating respiratory monitoring into routine care. In the future, our respiratory monitoring approach could be adapted into a more practical and comfortable form factor suitable for continuous non-invasive patient monitoring, with the potential to improve diagnosis, treatment, and ongoing disease management for patients with chronic respiratory illnesses.

8.3 Conclusion

Overall, the work presented in this thesis makes a compelling case for integrating respiratory monitoring into on-field sports performance tracking, highlighting its potential to enhance the understanding and assessment of athlete exertion in real-time, high-intensity environments. Breathing audio collected via a microphone-instrumented mouthguard was proven to be able to capture not only the rate but also the volume of breathing. With continued advancements in sensor technology and the refinement of detection models through large-scale data approaches, there is strong potential for unobtrusive respiratory monitoring to become a standard practice in team sports — enabling deeper insights into sports physiology and improve player management for better health outcomes.

Appendix A

Common Head Impact Severity Measures

Table A.1: A list of some of the commonly accepted kinematics-based measures of head impact severity, where a represents linear acceleration, α represents angular acceleration, ω represents angular velocity, and t represents time. In GAMBIT, n , m , and s are assigned constants and are often taken to be 1. In CP, are regression coefficients with i indicating the index number. In HIP, m stands for mass and I stands for moment of inertia. In HITsp, is the impact location coefficient, and In PRHIC HIP_rot is the rotational component of HIP.

Name	Measure
Peak LA	a_{peak}

Peak AA

$$\alpha_{peak}$$

GSI

$$I = \int_{\Delta t} a(t)^{2.5} dt$$

HIC

$$HIC = \left[\frac{\int_{t_1}^{t_2} a(t) dt}{(t_2 - t_1)} \right]^{2.5} (t_2 - t_1)$$

GAMBIT

$$G(t) = \left[\left(\frac{a(t)}{a_c} \right)^n + \left(\frac{\alpha(t)}{\alpha_c} \right)^m \right]^{1/s}$$

KLC

$$KLC = 0.004718 \cdot \omega_r + 0.000224 \cdot HIC$$

CP

$$CP = \frac{1}{1 + e^{-(\beta_0 + \beta_1 a + \beta_2 \alpha + \beta_3 a \alpha)}}$$

HIP

$$HIP = m \sum_i a_i \int a_i dt + \sum_j I_{jj} \alpha_j \int \alpha_j dt$$

$$\text{HIT}_{\text{sp}} \quad wPCS = x_n \cdot 10 \cdot ([0.4718 \cdot sGSI + 0.4742 \cdot sHIC + 0.4336 \cdot sLA + 0.2164 \cdot sAA] + 2))$$

$$\text{BRIC}$$

$$BRIC = \frac{\omega_{max}}{\omega_{cr}} + \frac{\alpha_{max}}{\alpha_{cr}}$$

$$\text{BrIC}$$

$$BrIC = \sqrt{\left(\frac{\omega_x}{\omega_{xc}}\right)^2 + \left(\frac{\omega_y}{\omega_{yc}}\right)^2 + \left(\frac{\omega_z}{\omega_{zc}}\right)^2}$$

$$\text{RIC}$$

$$RIC = \left[\frac{\int_{t_1}^{t_2} \alpha(t) dt}{(t_2 - t_1)} \right]^{2.5} (t_2 - t_1)$$

$$\text{PRHIC}$$

$$PRHIC = \left[\frac{\int_{t_1}^{t_2} HIP_{rot} dt}{(t_2 - t_1)} \right]^{2.5} (t_2 - t_1)$$

Appendix B

Breathing Rate Logging Programs

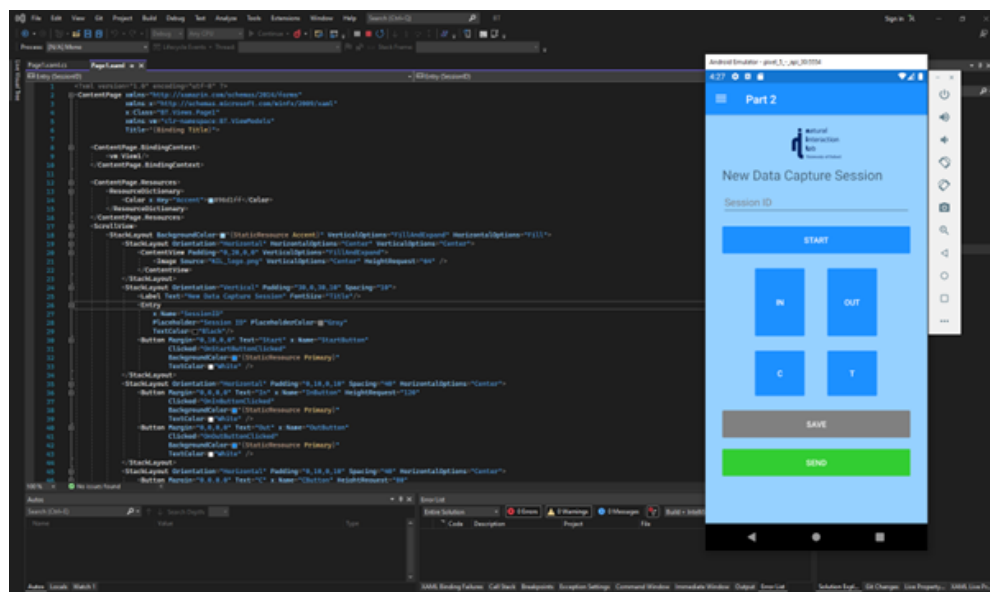
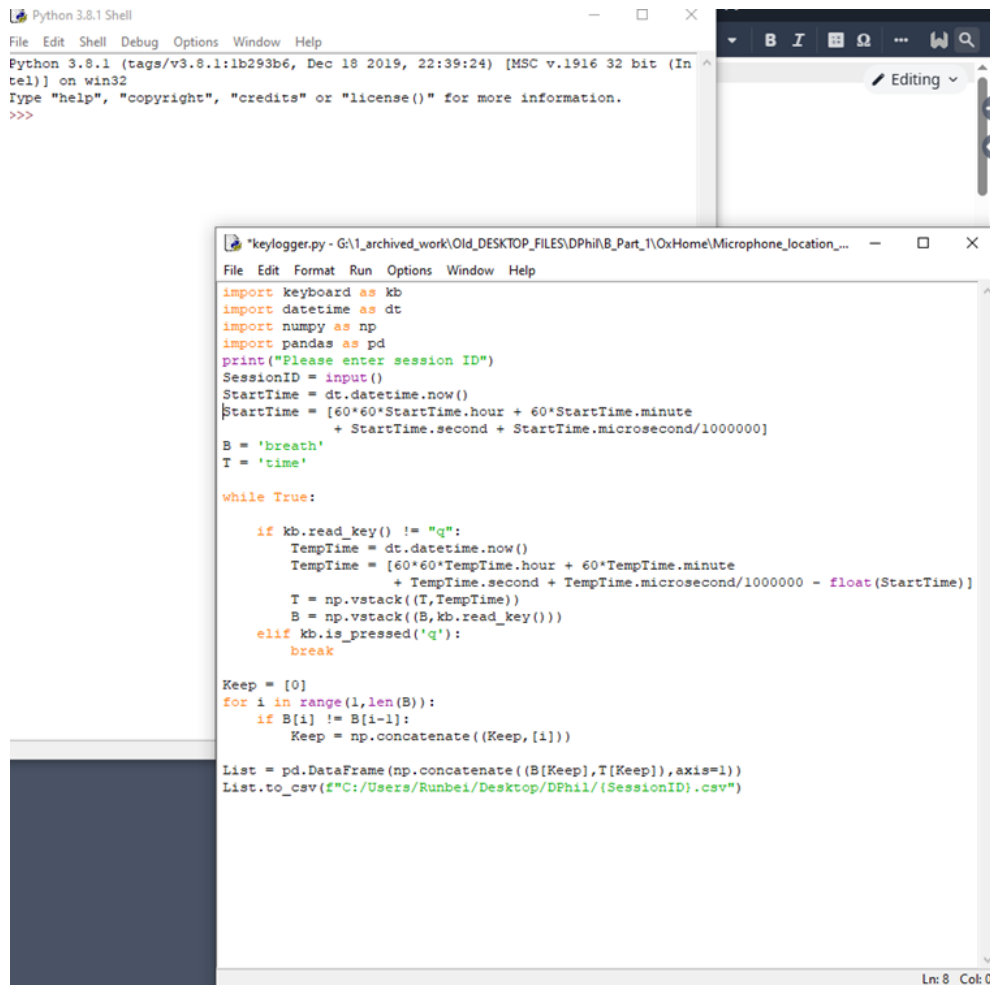


Figure B.1: A screen shot of the breath logging app running in an Android emulator.

An Android app (shown in figure B.2) was made in C# to log breaths during data captures where the participant is at rest, which is then used as the ground truth during data analyses. The app has an editable text field where the participant would enter the session ID. Two buttons, labelled “In” and “Out”, can be seen in the middle of the screen, for labeling inhales and exhales. When these buttons are pressed, they

register a timestamp. Once the data capture session has been completed, the data is sent to the researchers via e-mail.

A key-logger serving the same purpose was created in python for data capture sessions where a PC instead of a smartphone was used to log the breathing data.



The image shows two overlapping windows. The top window is a 'Python 3.8.1 Shell' with a menu bar (File, Edit, Shell, Debug, Options, Window, Help) and a command prompt showing the Python version and a prompt '(>>>'. The bottom window is a text editor titled 'keylogger.py' with a menu bar (File, Edit, Format, Run, Options, Window, Help). It contains the following Python code:

```
import keyboard as kb
import datetime as dt
import numpy as np
import pandas as pd
print("Please enter session ID")
SessionID = input()
StartTime = dt.datetime.now()
StartTime = [60*60*StartTime.hour + 60*StartTime.minute
             + StartTime.second + StartTime.microsecond/1000000]
B = 'breath'
T = 'time'

while True:
    if kb.read_key() != "q":
        TempTime = dt.datetime.now()
        TempTime = [60*60*TempTime.hour + 60*TempTime.minute
                    + TempTime.second + TempTime.microsecond/1000000 - float(StartTime)]
        T = np.vstack((T,TempTime))
        B = np.vstack((B,kb.read_key()))
    elif kb.is_pressed('q'):
        break

Keep = [0]
for i in range(1,len(B)):
    if B[i] != B[i-1]:
        Keep = np.concatenate((Keep,[i]))

List = pd.DataFrame(np.concatenate((B[Keep],T[Keep]),axis=1))
List.to_csv(f"C:/Users/Runbel/Desktop/DPhil/(SessionID).csv")
```

The status bar at the bottom right of the text editor shows 'Ln: 8 Col: 0'.

Figure B.2: The key-logger program running in python.

Appendix C

Large Dataset Metadata Tables

Here we have a detailed list of metadata on the large dataset described in Chapter 4. The dataset along with code to load and visualize the data has been uploaded to Github: <https://github.com/Oxford-NIL/YoYoTest>, and has been submitted to physionet.org for review.

ID	Gender	Age	Height (cm)	Weight (kg)	P (mmHg)	TA (°C)	RH (%)
D01	Female	24	170	73	764	18	44
D02	Female	25	150	50	764	18	49
D03	Female	19	158	52	767	18	56
D04	Female	22	168	54	764	17	50
D05	Male	20	182	76	766	16	45
D06	Male	18	194	92	760	19	36
D07	Female	23	182	75	760	23	27
D08	Male	22	182	63	760	26	29
D09	Male	18	184	80	765	16	40
D10	Male	33	170	70	767	19	54
D11	Male	23	185	82.5	765	13	50
D13	Female	24	169	78.5	767	16	57
D14	Female	27	160	50	761	14	41
D15	Male	25	170	80	748	19	38
D16	Male	20	176	65	760	21	38
D17	Male	28	170	75	764	17	51
D18	Male	24	177	70	767	18	54
D19	Female	24	160	60	755	31	47
D20	Male	43	179	71	764	17	61
D21	Female	40	164	87	756	36	40
D24	Female	26	174	75	756	32	58

Table C.1: First 8 columns from the first subset’s metadata file, including participants’ information and environmental measures during the data capture session. P stands for Barometric Pressure (mmHg), TA stands for Ambient Temperature (°C), and RH stands for Ambient Relative Humidity (%).

ID	COSMED File Name	HR Included?	YoYo Start Time (s)	YoYo Start Index	YoYo End Time (s)	YoYo End Index
D01	D01_COSMED_1.csv	1	1063	275	1206	357
D02	D02_COSMED_1.csv	1	1225	481	1517	732
D03	D03_COSMED_1.csv	1	759	193	1003	348
D04	D04_COSMED_1.csv	1	1149	317	1429	545
D05	D05_COSMED_1.csv	0	1003	167	1600	650
D06	D06_COSMED_1.csv	1	853	169	1708	789
D07	D07_COSMED_1.csv	1	1264	438	1561	648
D08	D08_COSMED_1.csv	1	908	267	1823	1001
D09	D09_COSMED_1.csv	0	1254	329	1852	795
D10	D10_COSMED_1.csv	1	1444	347	1972	628
D11	D11_COSMED_1.csv	0	847	245	1786	990
D13	D13_COSMED_1.csv	1	738	284	939	424
D14	D14_COSMED_1.csv	1	1348	345	1544	481
D15	D15_COSMED_1.csv	1	659	215	839	359
D16	D16_COSMED_1.csv	1	1018	222	1585	679
D17	D17_COSMED_1.csv	1	1162	301	1416	425
D18	D18_COSMED_1.csv	1	674	255	828	364
D19	D19_COSMED_1.csv	1	498	148	672	294
D20	D20_COSMED_1.csv	1	700	209	1314	682
D21	D21_COSMED_1.csv	1	988	270	1087	348
D24	D24_COSMED_1.csv	1	1157	357	1510	519

Table C.2: COSMED data information columns from the first subset’s meta-data file. The ‘HR Included?’ column is a binary indicator of whether heart rate data is included, with 1 meaning yes, and 0 meaning no.

ID	RPE File Name	Session RPE	YoYo Test Level	Audio Included?	Audio File Name	Audio to COSMED Alignment (s)
D01	D01_RPE_1.csv	8	12.2	1	D01_Audio_1.wav	1.7
D02	D02_RPE_1.csv	8	14.2	0		
D03	D03_RPE_1.csv	5	13.4	0		
D04	D04_RPE_1.csv	5	14.2	1	D04_Audio_1.wav	2.8
D05	D05_RPE_1.csv	7	16.2	1	D05_Audio_1.wav	2.0
D06	D06_RPE_1.csv	7	17.8	1	D06_Audio_1.wav	2.7
D07	D07_RPE_1.csv	4	14.3	1	D07_Audio_1.wav	0.4
D08	D08_RPE_1.csv	8	18.1	1	D08_Audio_1.wav	2.4
D09	D09_RPE_1.csv	7	16.2	1	D09_Audio_1.wav	1.0
D10	D10_RPE_1.csv	7	14.2	1	D10_Audio_1.wav	-2.8
D11	D11_RPE_1.csv	7	18.3	1	D11_Audio_1.wav	-2.9
D13	D13_RPE_1.csv	7	13.2	1	D13_Audio_1.wav	0.7
D14	D14_RPE_1.csv	5	13.2	1	D14_Audio_1.wav	-1.2
D15	D15_RPE_1.csv	7	13.1	1	D15_Audio_1.wav	0.6
D16	D16_RPE_1.csv	8	15.8	1	D16_Audio_1.wav	-1.2
D17	D17_RPE_1.csv	4	13.3	0		
D18	D18_RPE_1.csv	7	12.3	0		
D19	D19_RPE_1.csv	5	13.1	1	D19_Audio_1.wav	-0.2
D20	D20_RPE_1.csv	7	16.2	1	D20_Audio_1.wav	-1.5
D21	D21_RPE_1.csv	6	11.2	1	D21_Audio_1.wav	0.6
D24	D24_RPE_1.csv	7	11.2	1	D24_Audio_1.wav	-1.5

Table C.3: RPE and audio data information columns from the first subset’s metadata file. The session RPE was evaluated 30 minutes after the end of the Yo-Yo test, rated on Borg’s CR10 RPE scale. The ‘Audio Included?’ column is a binary indicator of whether audio data is included, with 1 meaning yes, and 0 meaning no.

ID	Gender	Age	Height (cm)	Weight (kg)	P (mmHg)	TA (°C)	RH (%)
D01	Female	24	170	73	765	18	53
D02	Female	25	150	50	766	16	59
D03	Female	19	158	52	765	23	42
D04	Female	22	168	54	765	21	51
D05	Male	20	182	76	755	32	53
D06	Male	19	194	92	766	17	58
D08	Male	22	182	63	765	18	47
D09	Male	18	184	80	767	14	60
D10	Male	33	170	70	765	20	48
D11	Male	23	185	82.5	767	16	59
D13	Female	24	169	78.5	765	21	49
D14	Female	27	160	50	757	31	52
D15	Male	25	170	80	764	18	56
D16	Male	20	176	65	757	30	55
D17	Male	29	170	75	765	20	53
D18	Male	24	177	70	765	22	48
D20	Male	43	179	71	756	30	56
D21	Female	40	164	87	757	25	64
D24	Female	26	174	75	757	28	55

Table C.4: First 8 columns from the second subset’s metadata file, including participants’ information and environmental measures during the data capture session. P stands for Barometric Pressure (mmHg), TA stands for Ambient Temperature (°C), and RH stands for Ambient Relative Humidity (%).

ID	COSMED File Name	HR Included?	YoYo Start Time (s)	YoYo Start Index	YoYo End Time (s)	YoYo End Index
D01	D01_COSMED_2.csv	1	1294	234	1442	323
D02	D02_COSMED_2.csv	1	526	193	839	417
D03	D03_COSMED_2.csv	1	381	147	677	357
D04	D04_COSMED_2.csv	1	832	320	1250	649
D05	D05_COSMED_2.csv	1	491	172	1292	842
D06	D06_COSMED_2.csv	1	508	151	1604	1029
D08	D08_COSMED_2.csv	1	615	226	1428	979
D09	D09_COSMED_2.csv	0	1147	347	1804	843
D10	D10_COSMED_2.csv	1	843	241	1375	539
D11	D11_COSMED_2.csv	1	1053	267	2264	1274
D13	D13_COSMED_2.csv	1	626	228	865	396
D14	D14_COSMED_2.csv	1	695	180	890	321
D15	D15_COSMED_2.csv	1	731	184	1023	430
D16	D16_COSMED_2.csv	1	715	197	1397	749
D17	D17_COSMED_2.csv	1	766	249	1027	381
D18	D18_COSMED_2.csv	1	934	211	1134	334
D20	D20_COSMED_2.csv	1	717	184	1311	654
D21	D21_COSMED_2.csv	1	705	187	871	318
D24	D24_COSMED_2.csv	1	487	159	607	233

Table C.5: COSMED data information columns from the second subset’s metadata file. The ‘HR Included?’ column is a binary indicator of whether heart rate data is included, with 1 meaning yes, and 0 meaning no.

ID	RPE File Name	Session RPE	YoYo Test Level	Audio Included?	Audio File Name	Audio to COSMED Alignment (s)
D01	D01_RPE_2.csv	9	12.2	0		
D02	D02_RPE_2.csv	8	14.3	1	D02_Audio_2.wav	0.3
D03	D03_RPE_2.csv	5	14.3	0		
D04	D04_RPE_2.csv	7	15.1	0		
D05	D05_RPE_2.csv	6	17.5	1	D05_Audio_2.wav	1.2
D06	D06_RPE_2.csv	8	19.4	1	D06_Audio_2.wav	-0.2
D08	D08_RPE_2.csv	9	17.5	1	D08_Audio_2.wav	1.3
D09	D09_RPE_2.csv	7	16.5	1	D09_Audio_2.wav	1.0
D10	D10_RPE_2.csv	7	14.2	1	D10_Audio_2.wav	-0.5
D11	D11_RPE_2.csv	8	20.1	1	D11_Audio_2.wav	-0.4
D13	D13_RPE_2.csv	7	13.4	1	D13_Audio_2.wav	1.3
D14	D14_RPE_2.csv	8	13.2	1	D14_Audio_2.wav	-1.5
D15	D15_RPE_2.csv	6	13.4	1	D15_Audio_2.wav	0
D16	D16_RPE_2.csv	8	16.7	1	D16_Audio_2.wav	-0.1
D17	D17_RPE_2.csv	7	14.1	0		
D18	D18_RPE_2.csv	7	13.2	0		
D20	D20_RPE_2.csv	7	16.2	1	D20_Audio_2.wav	-1.3
D21	D21_RPE_2.csv	7	12.3	0		
D24	D24_RPE_2.csv	6	12.1	1	D24_Audio_2.wav	-2.1

Table C.6: RPE and audio data information columns from the second subset’s metadata file. The session RPE was evaluated 30 minutes after the end of the Yo-Yo test, rated on Borg’s CR10 RPE scale. The ‘Audio Included?’ column is a binary indicator of whether audio data is included, with 1 meaning yes, and 0 meaning no.

Appendix D

Breathing Rate and Minute Ventilation Plots Per YoYo Test

Here we have breathing rate and minute ventilation both captured by the COSMED and measured from the audio data, plotted against each other over time, for each individual YoYo test ran. The dotted vertically lines on the breathing rate plots indicate the YoYo test audio cues (beeps). In these plots, it can be seen that the COSMED data for Subject D01 - Test1, Subject D02 - Test 2, and Subject 8 - Test 2 were particularly noisy.

Figure D.1: Subject D01 - Test 1

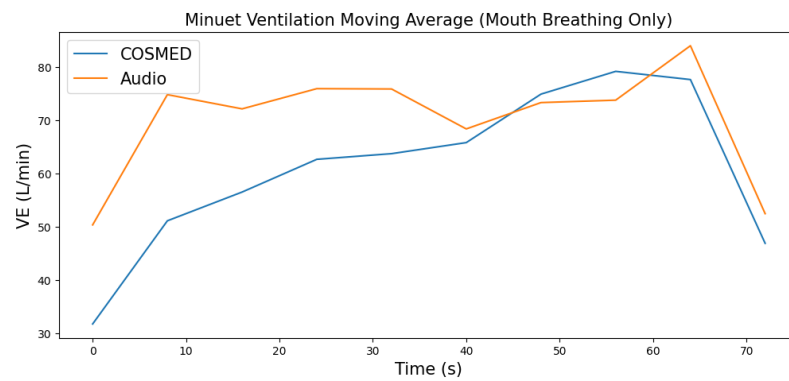
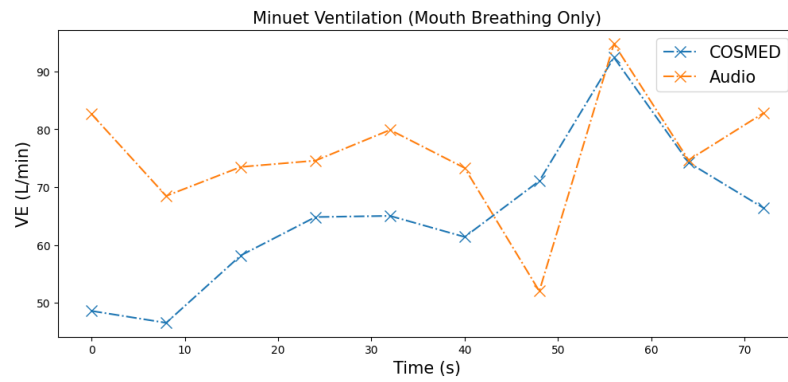
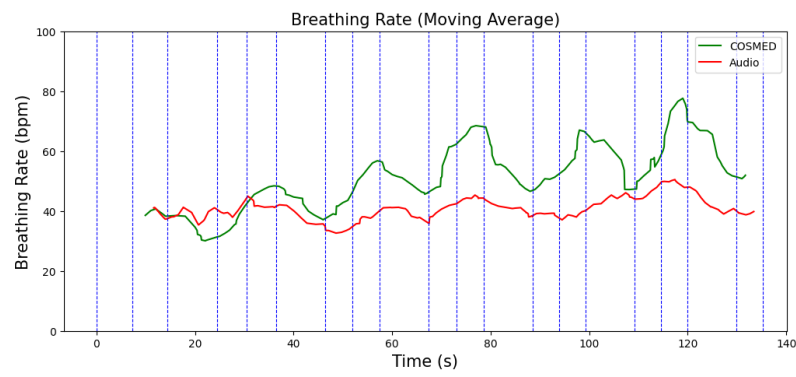
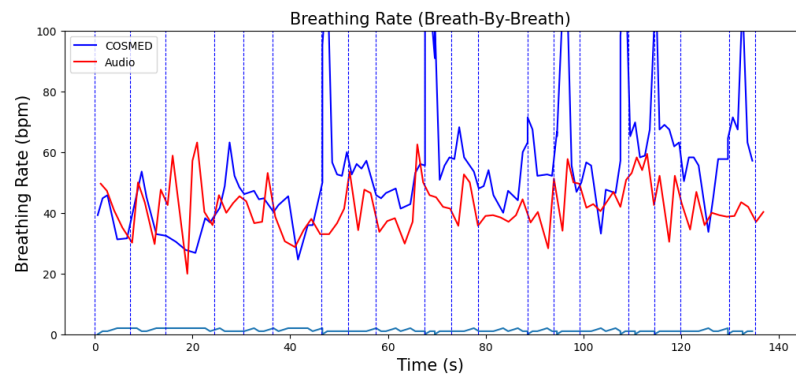


Figure D.2: Subject D02 - Test 2

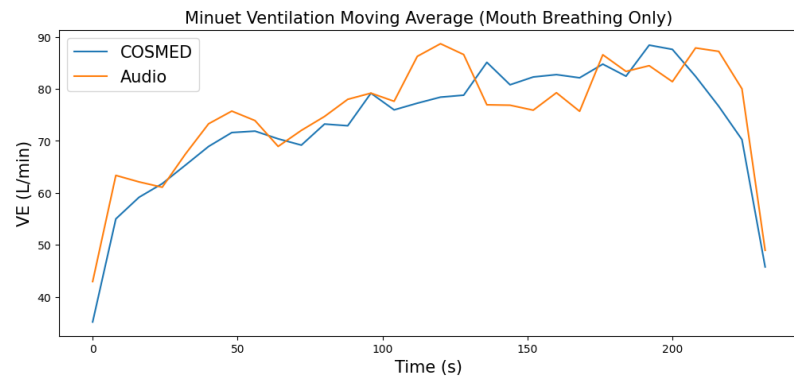
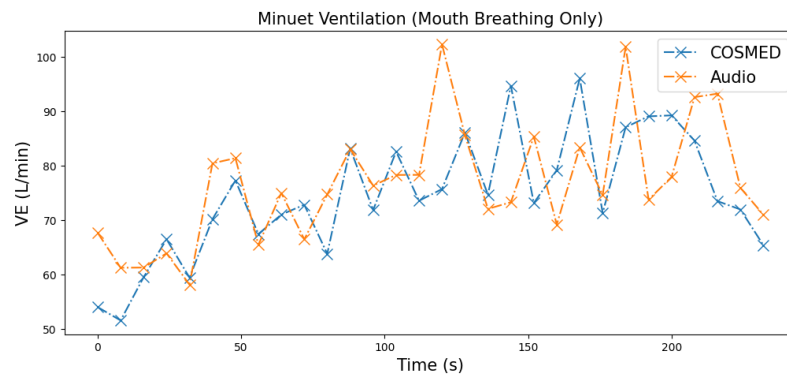
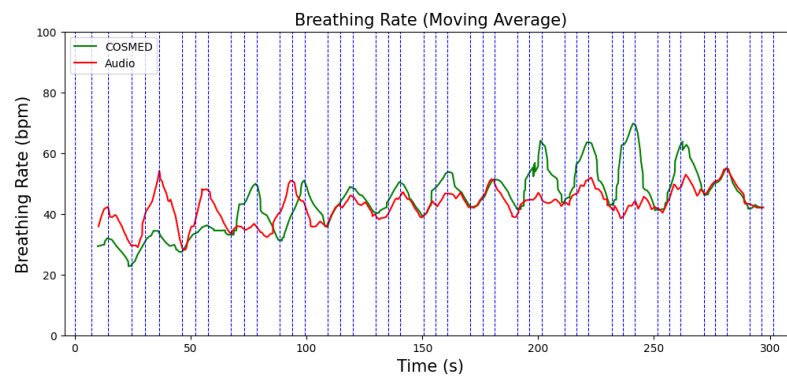
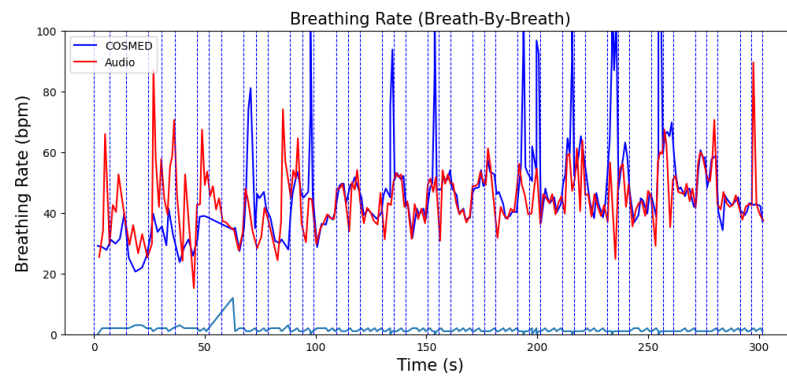


Figure D.3: Subject D04 - Test 1

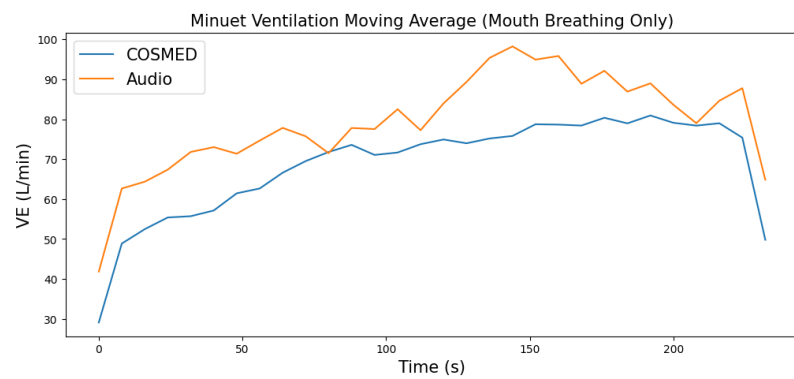
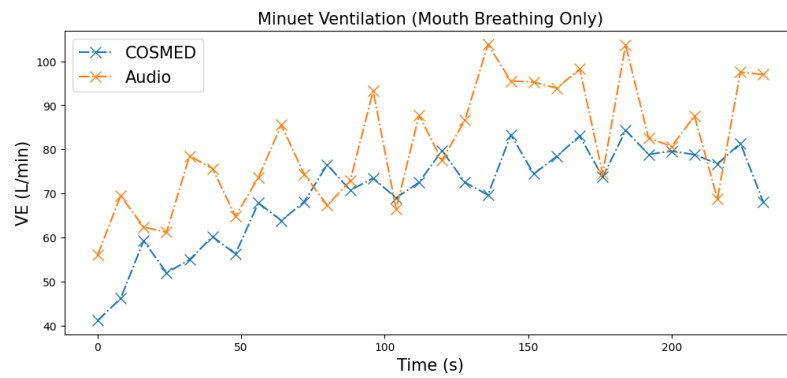
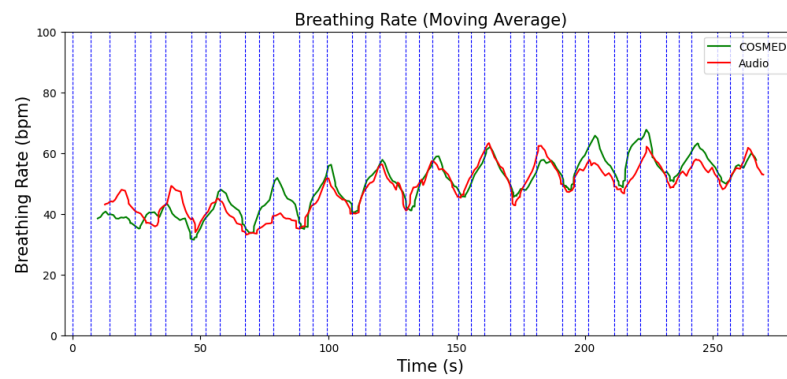
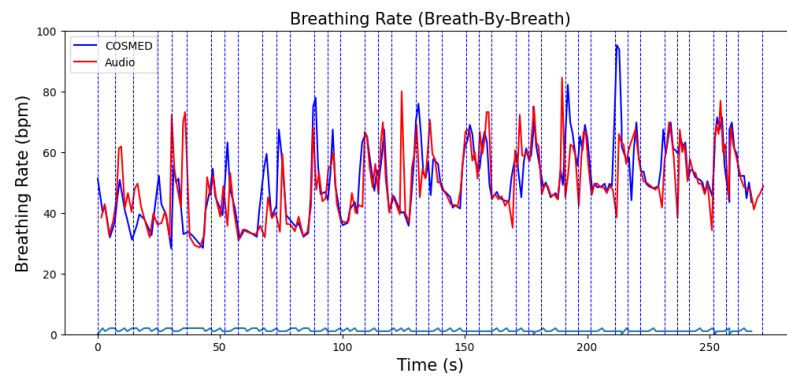


Figure D.4: Subject D05 - Test 1

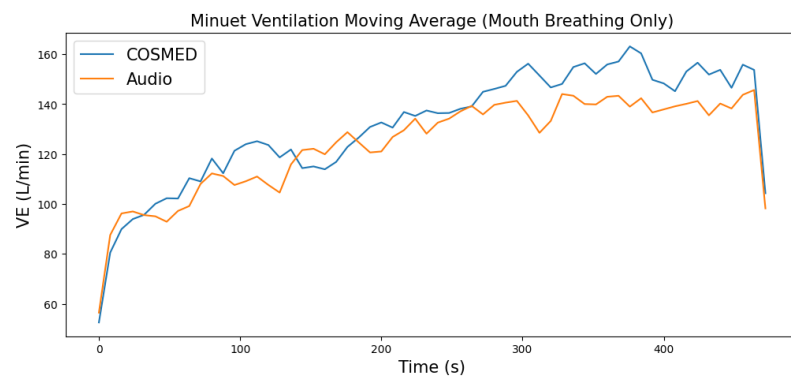
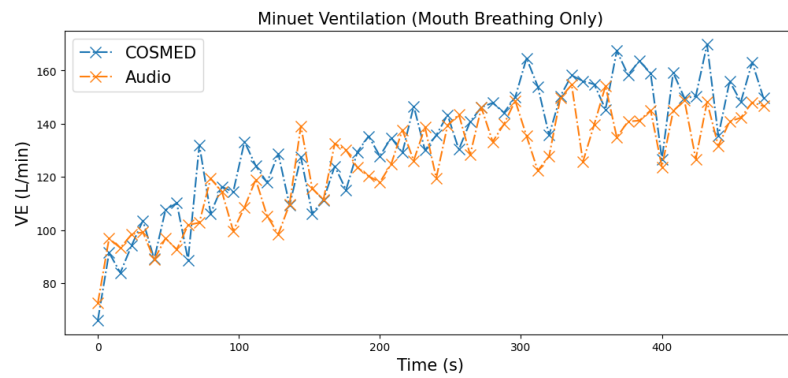
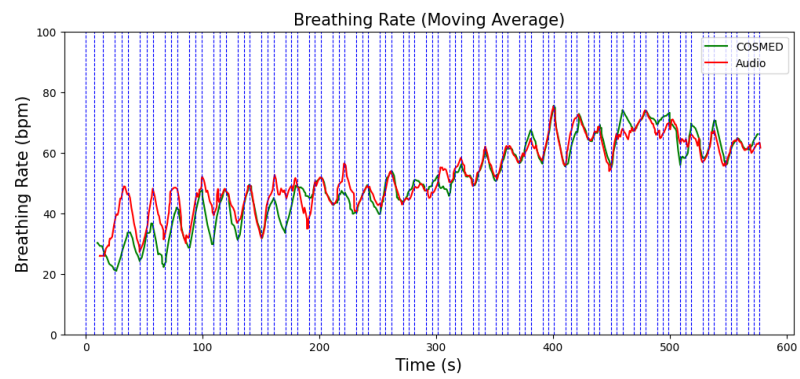
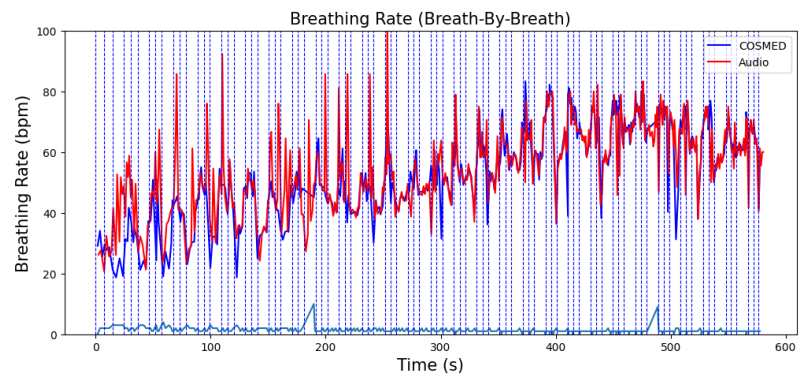


Figure D.5: Subject D05 - Test 2

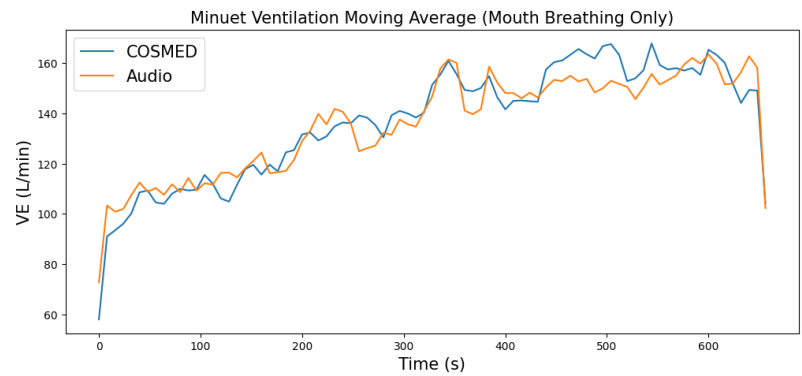
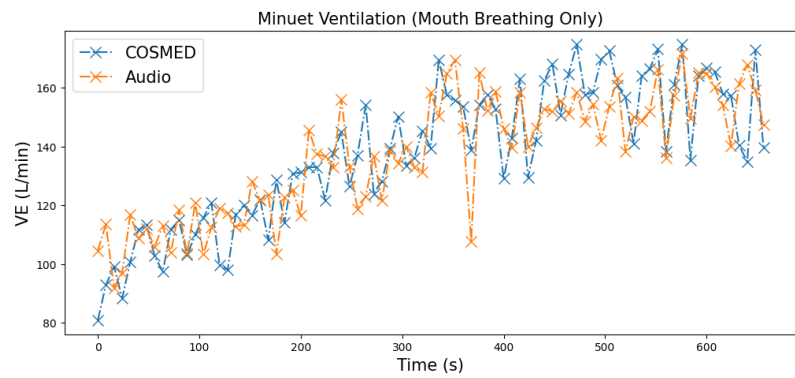
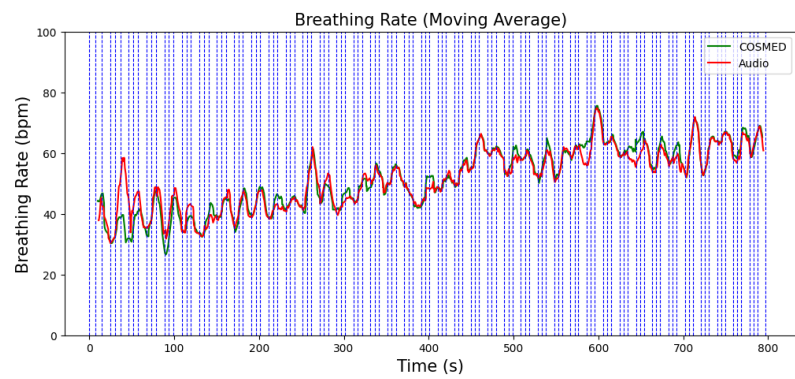
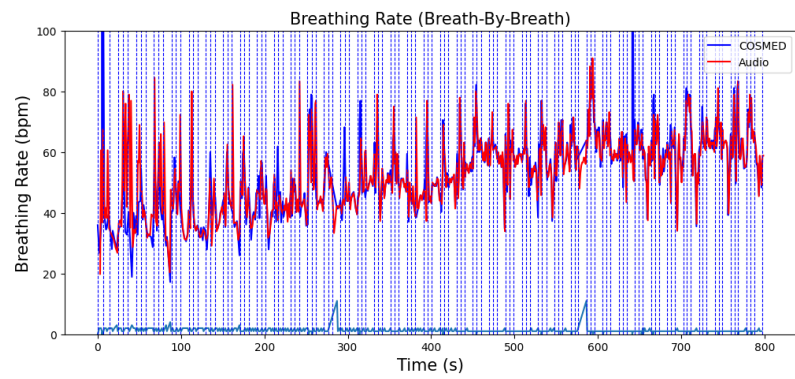


Figure D.6: Subject D06 - Test 1

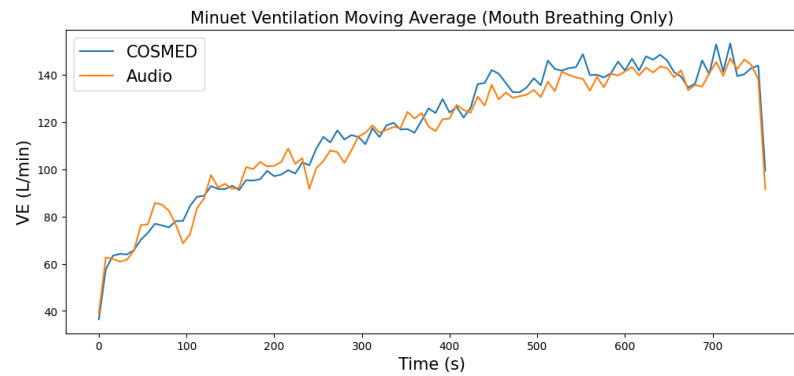
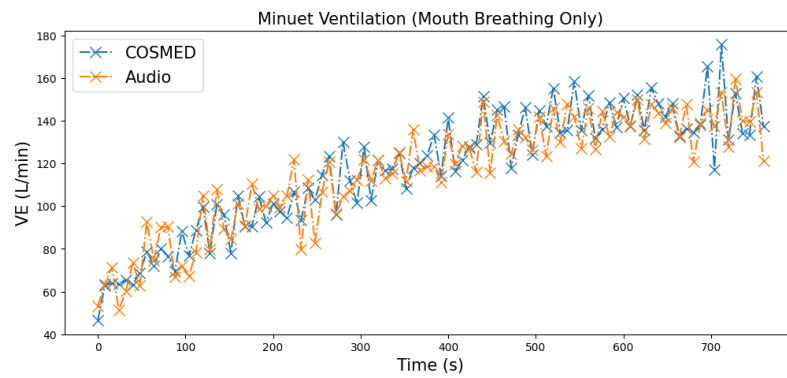
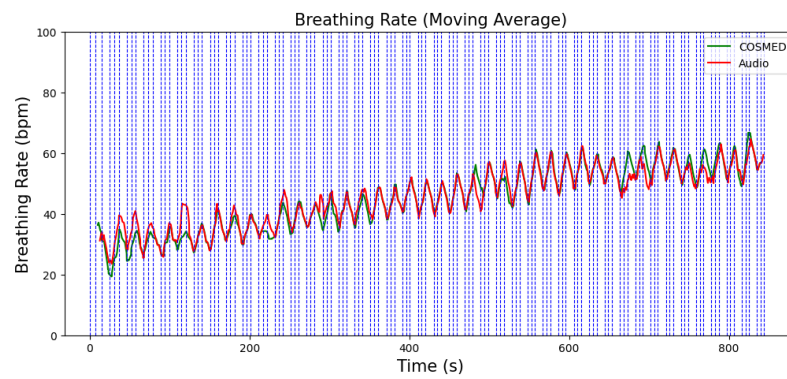
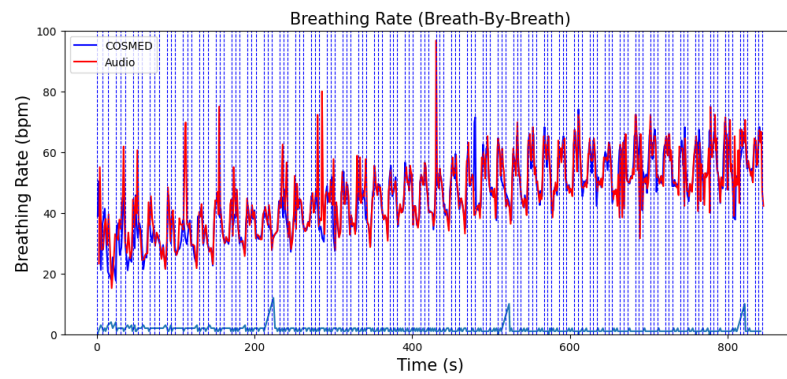


Figure D.7: Subject D06 - Test 2

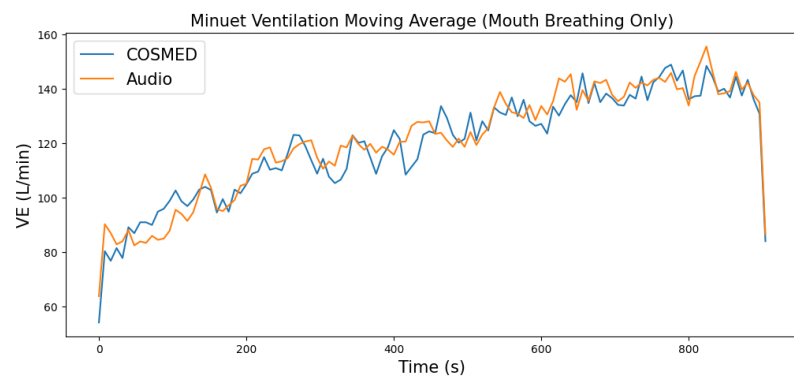
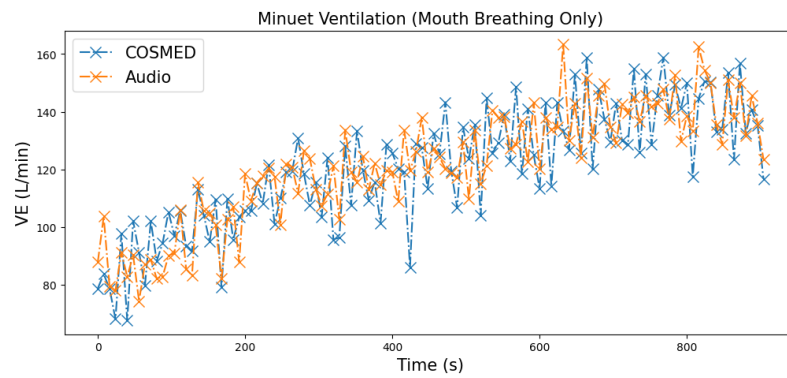
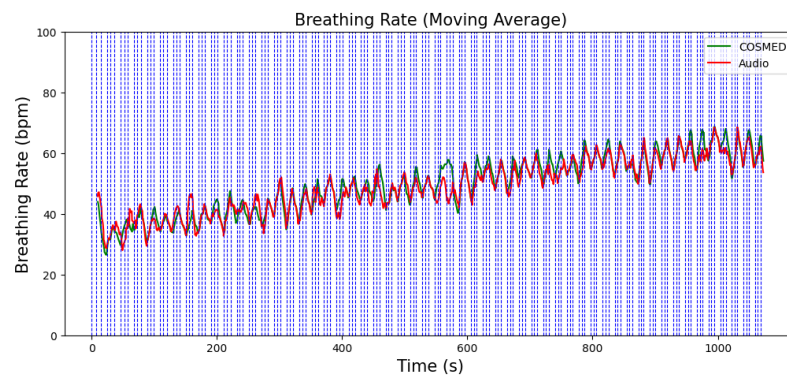
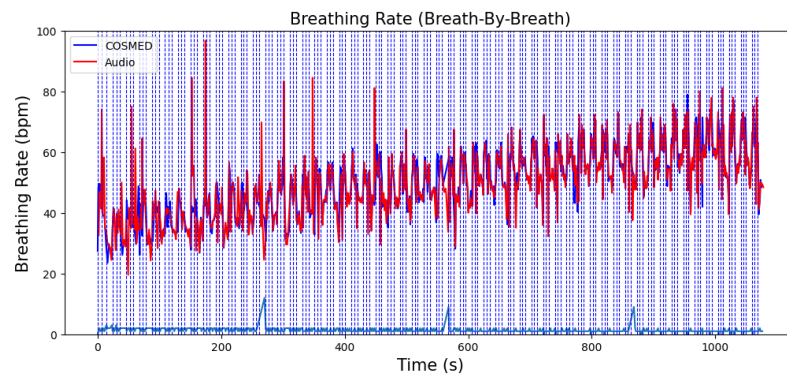


Figure D.8: Subject D07 - Test 1

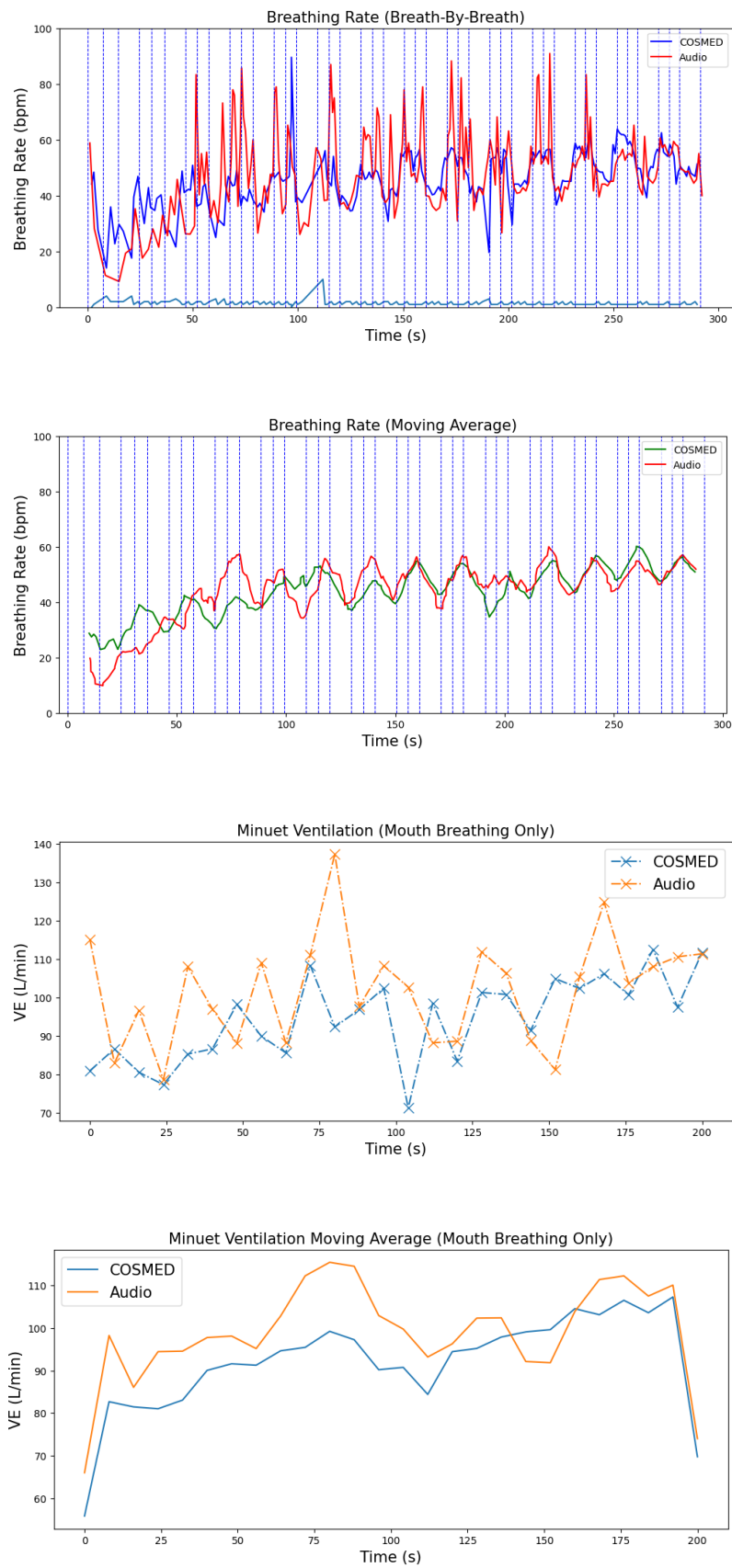


Figure D.9: Subject D08 - Test 1

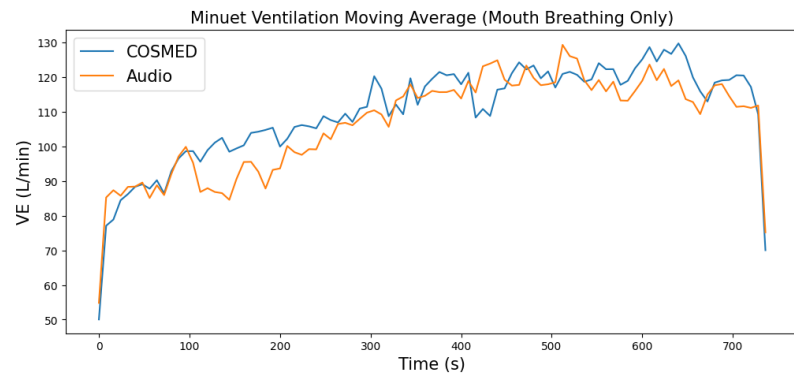
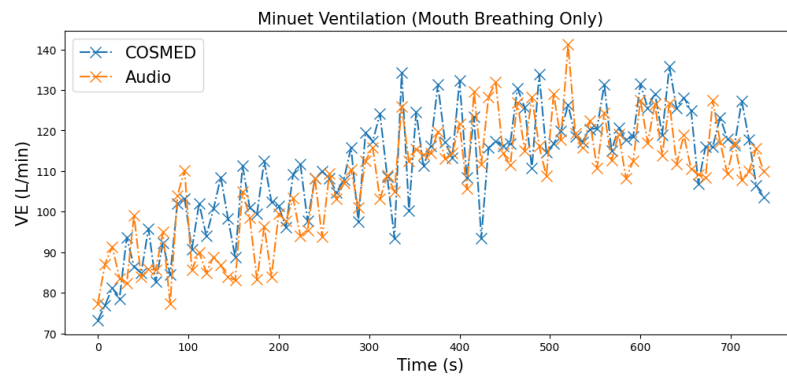
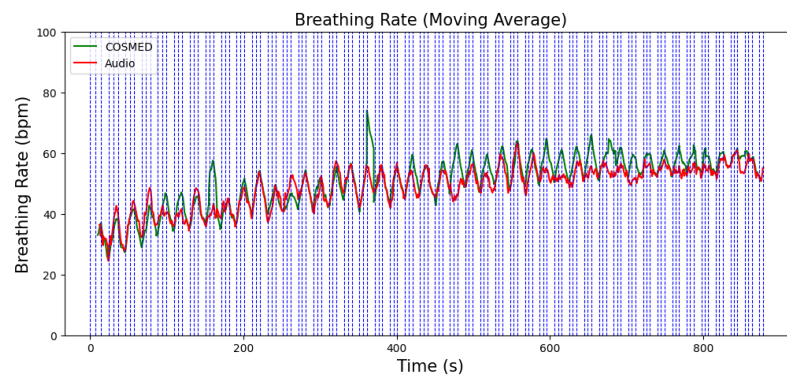
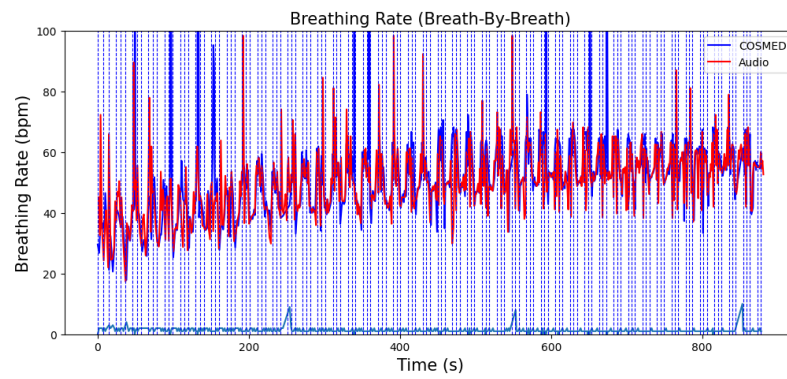


Figure D.10: Subject D08 - Test 2

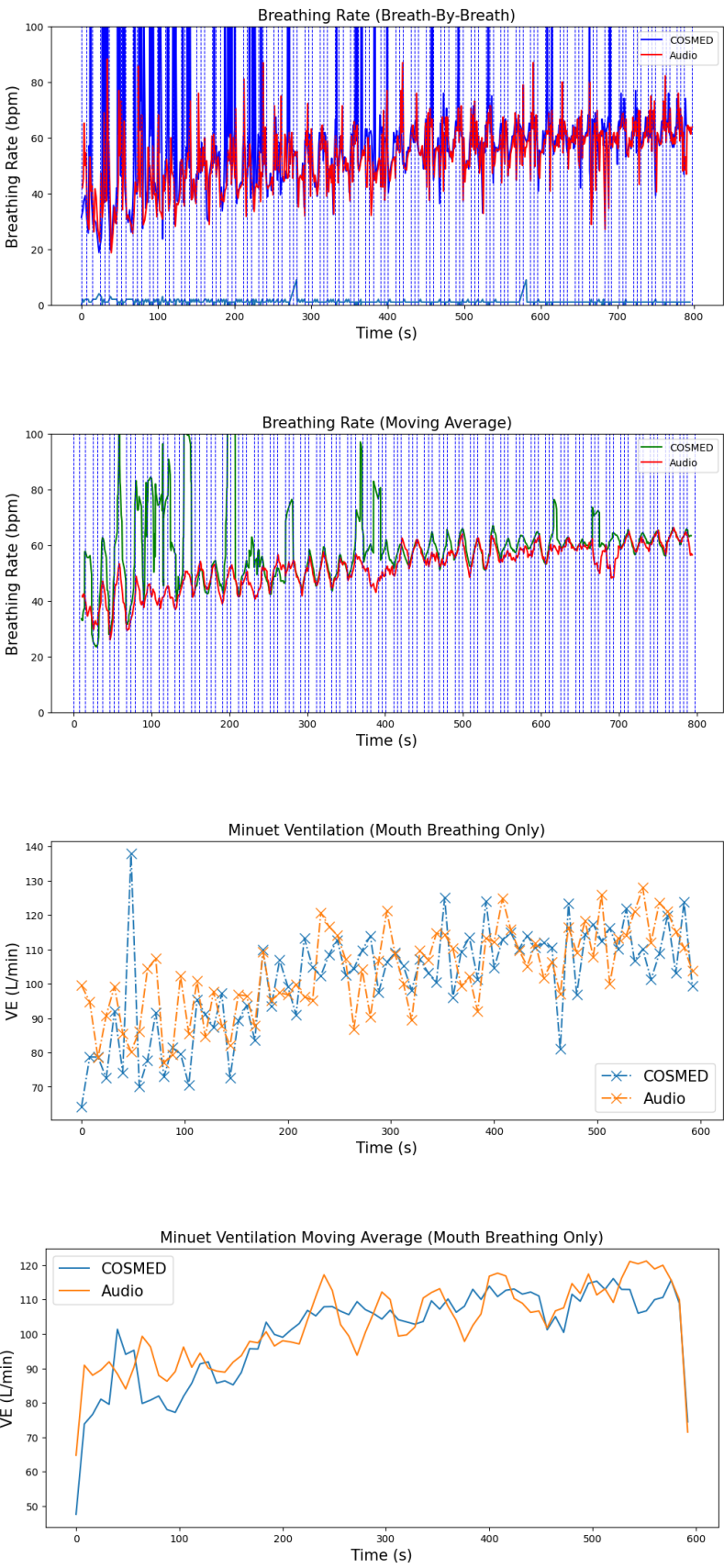


Figure D.11: Subject D09 - Test 1

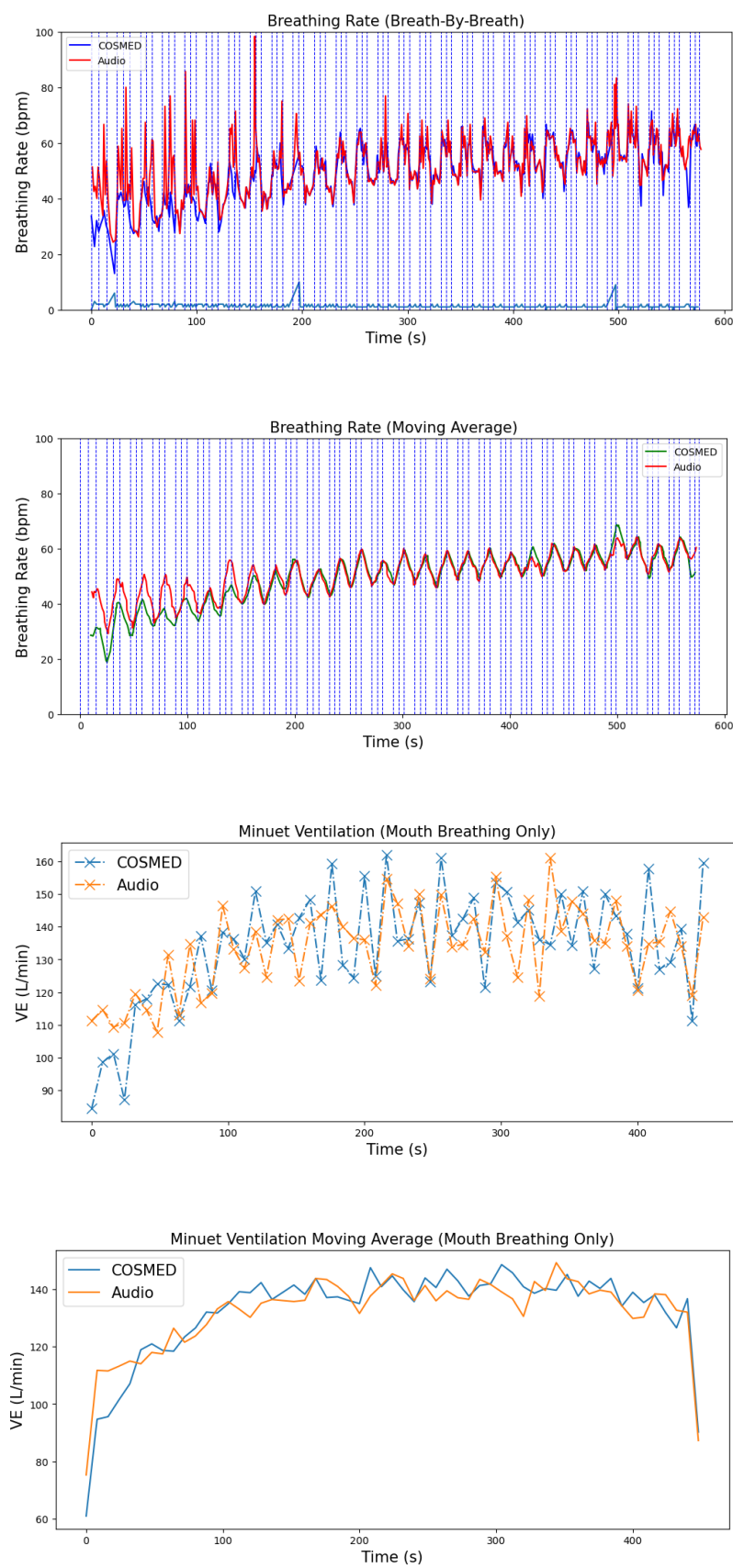


Figure D.12: Subject D09 - Test 2

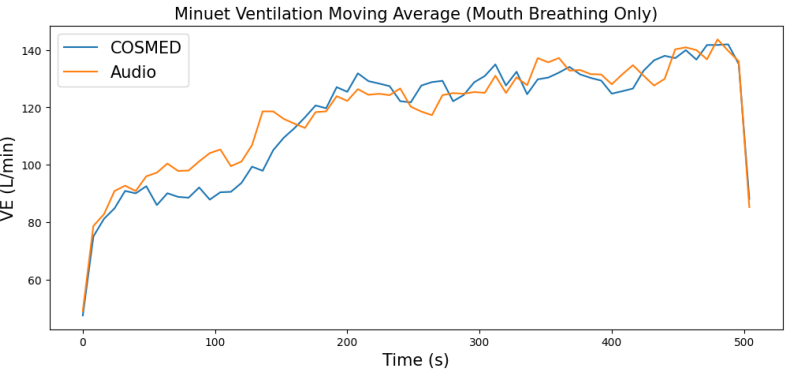
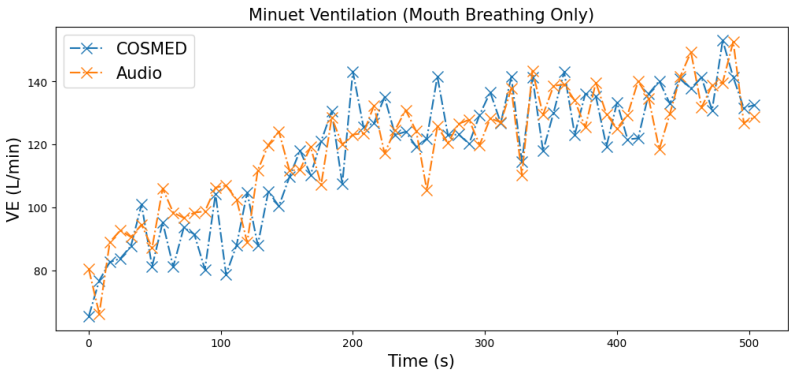
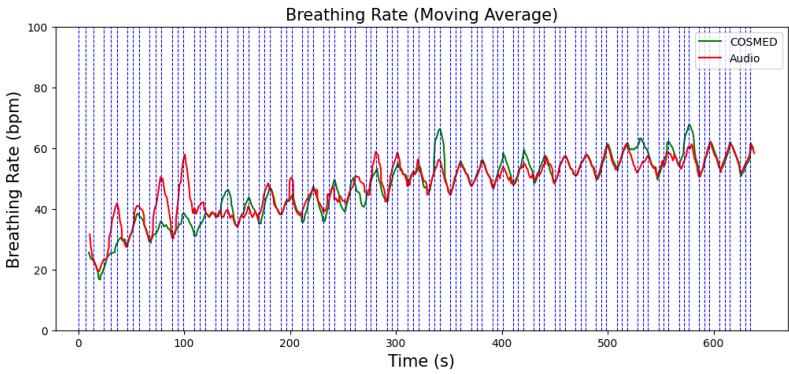
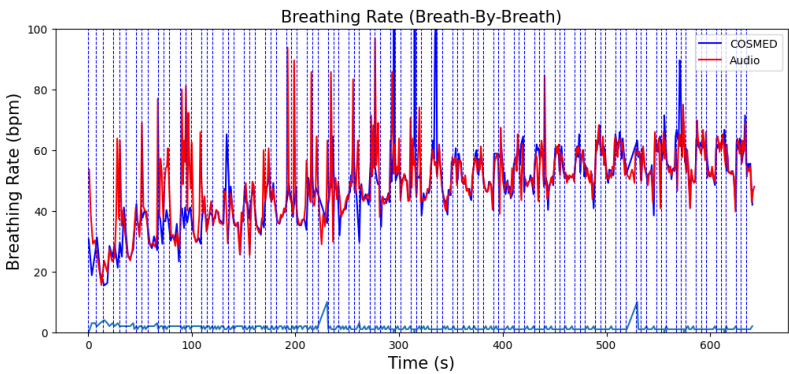


Figure D.13: Subject D10 - Test 1

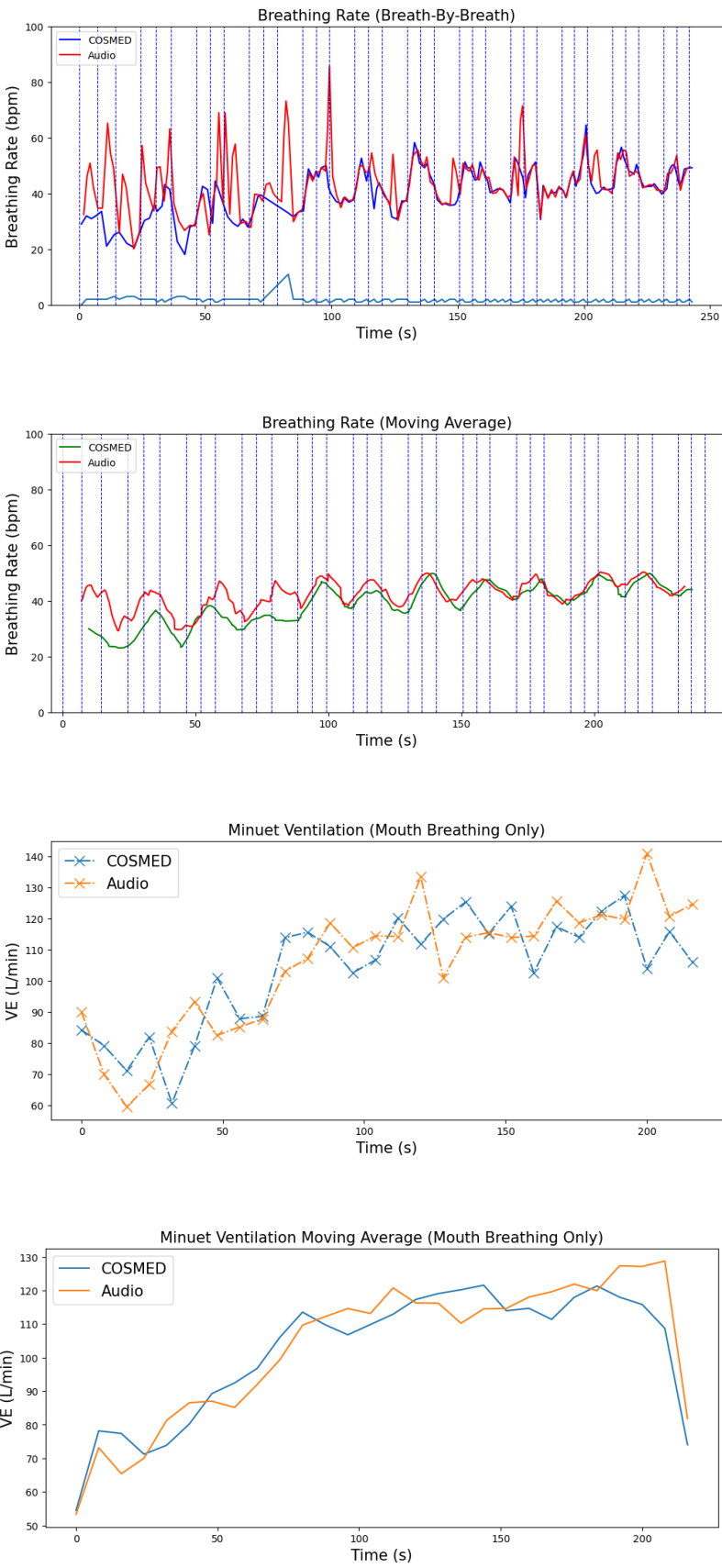


Figure D.14: Subject D10 - Test 2

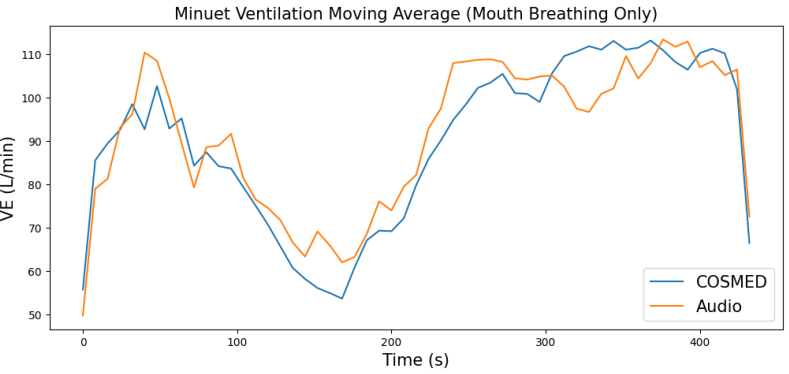
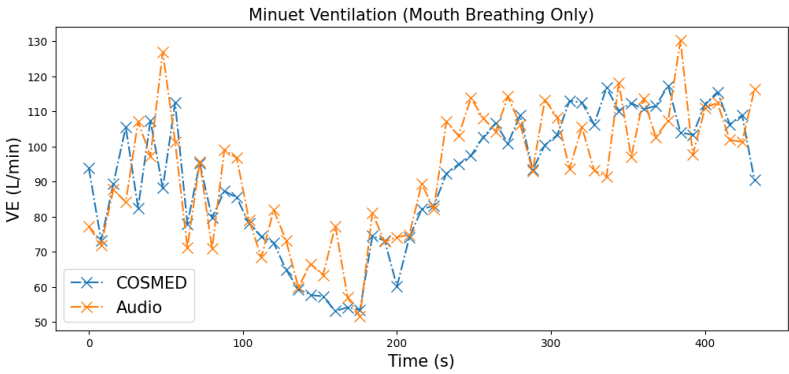
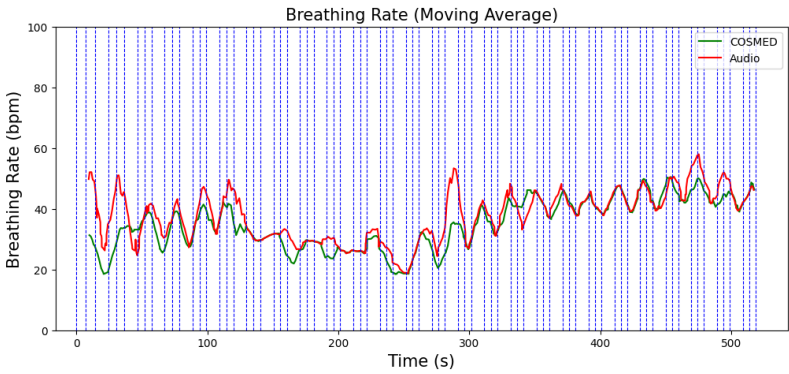
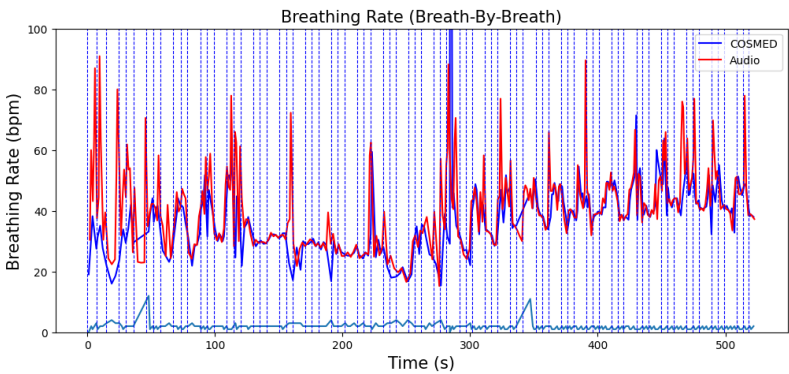


Figure D.15: Subject D11 - Test 1

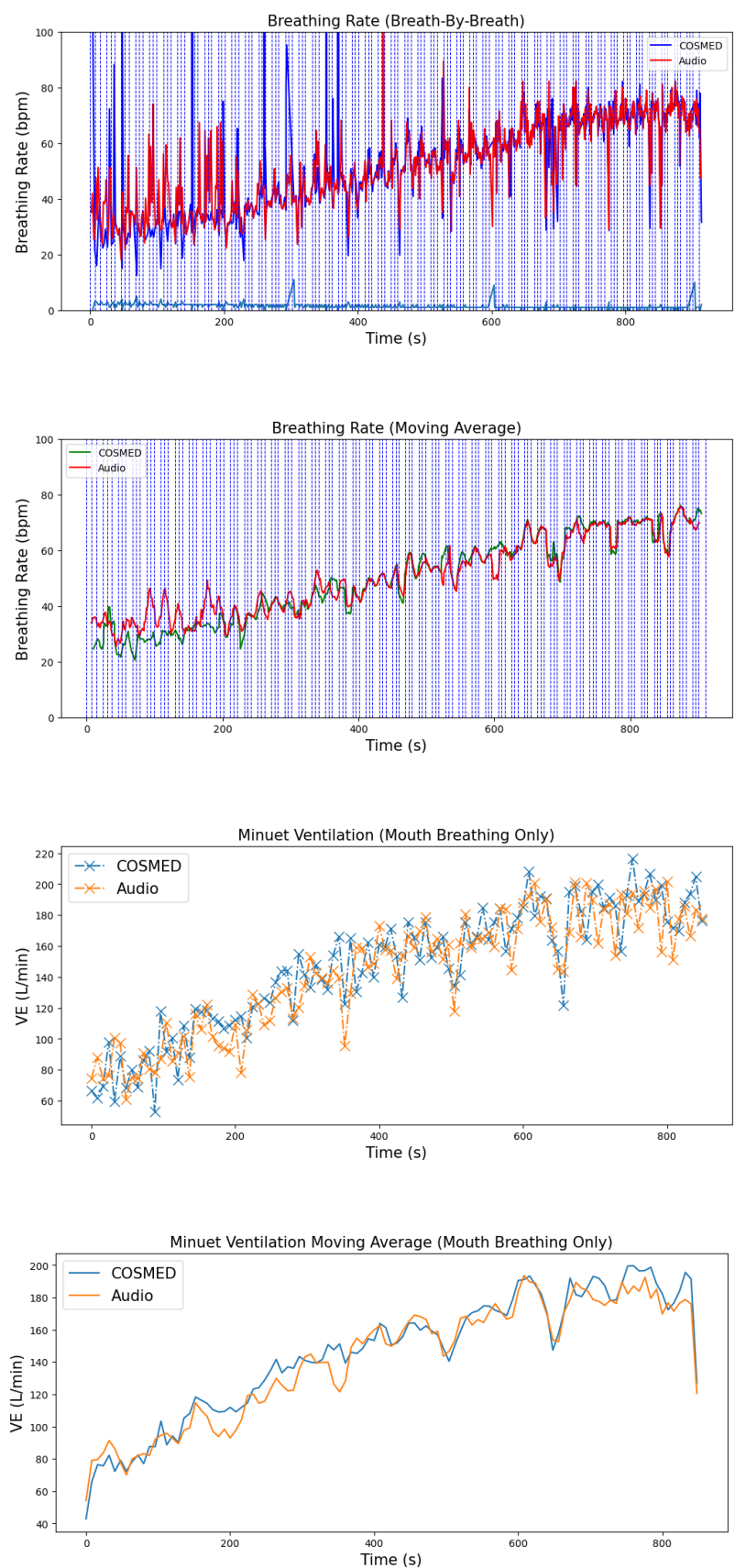


Figure D.16: Subject D11 - Test 2

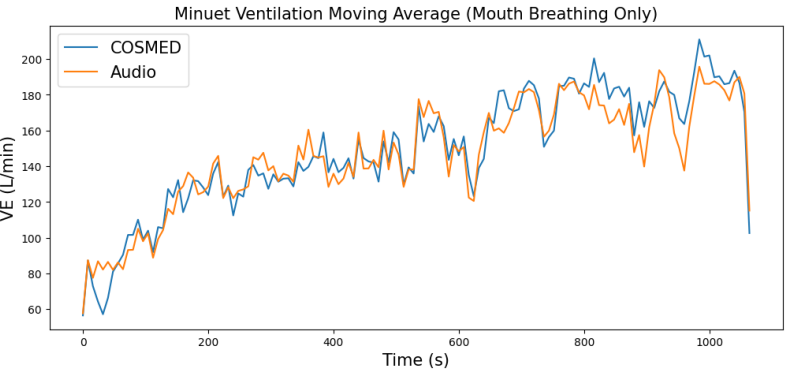
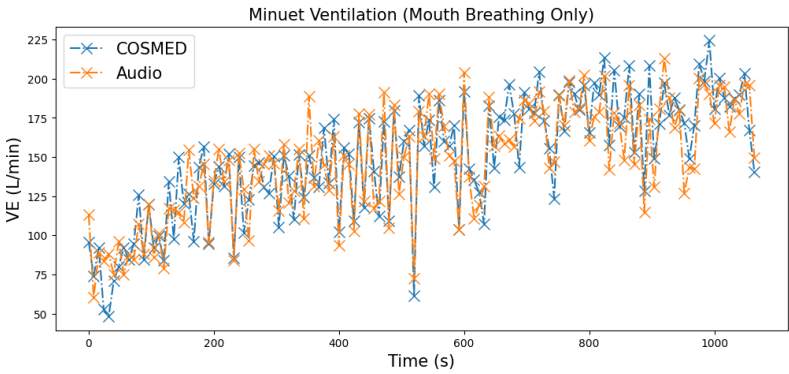
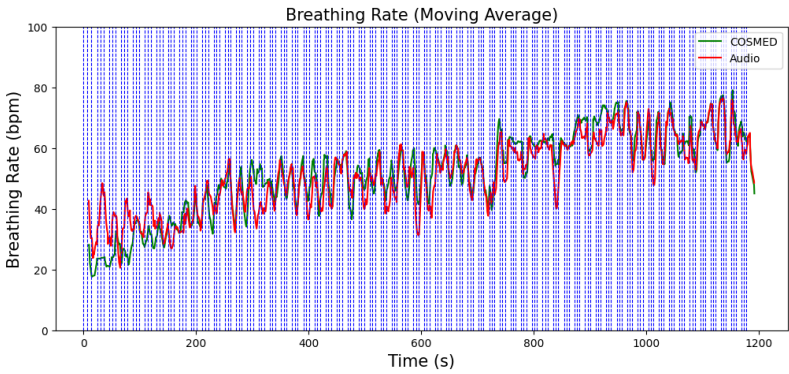
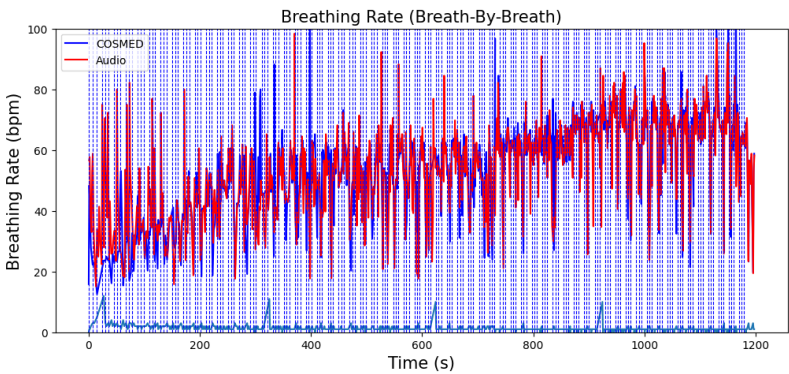


Figure D.17: Subject D13 - Test 1

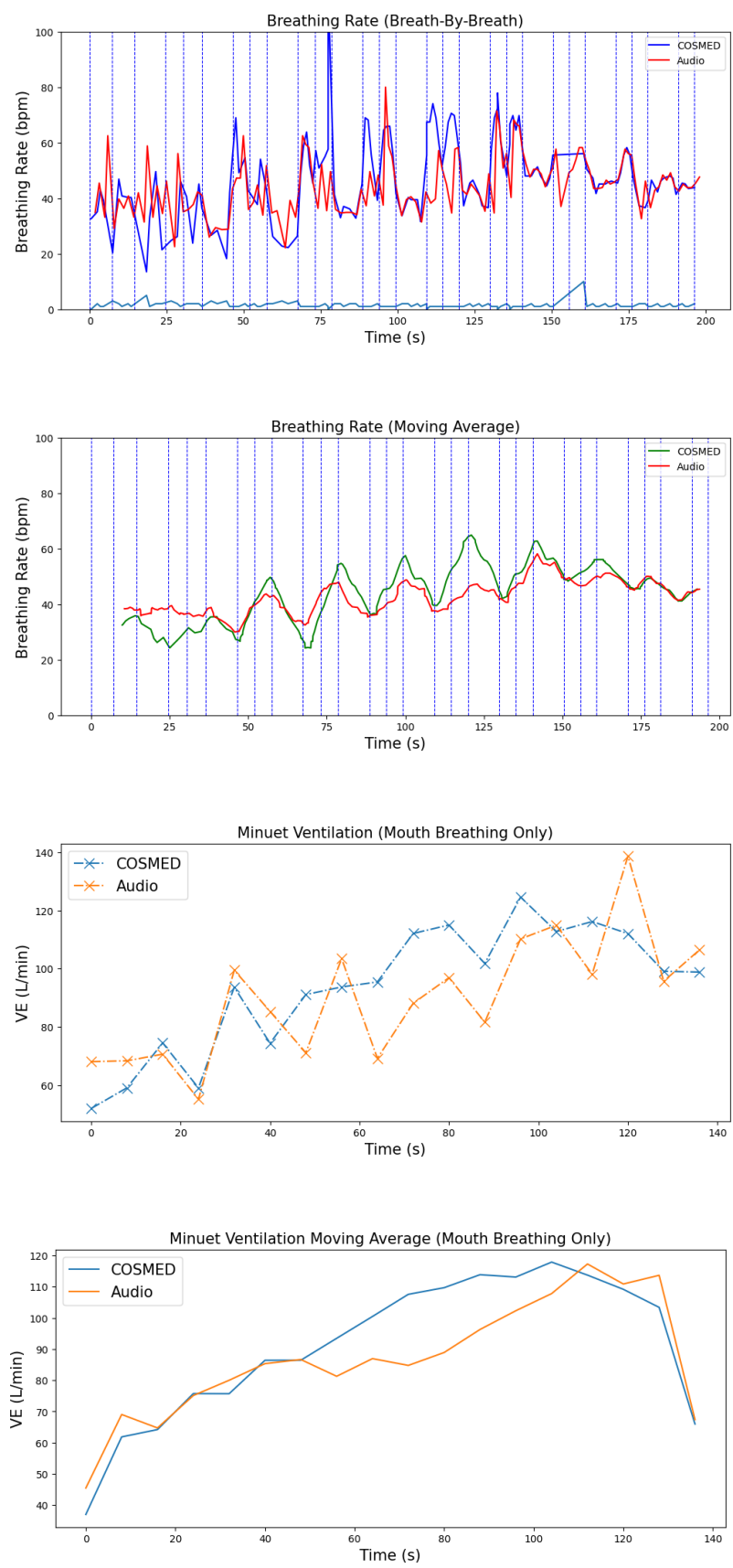


Figure D.18: Subject D13 - Test 2

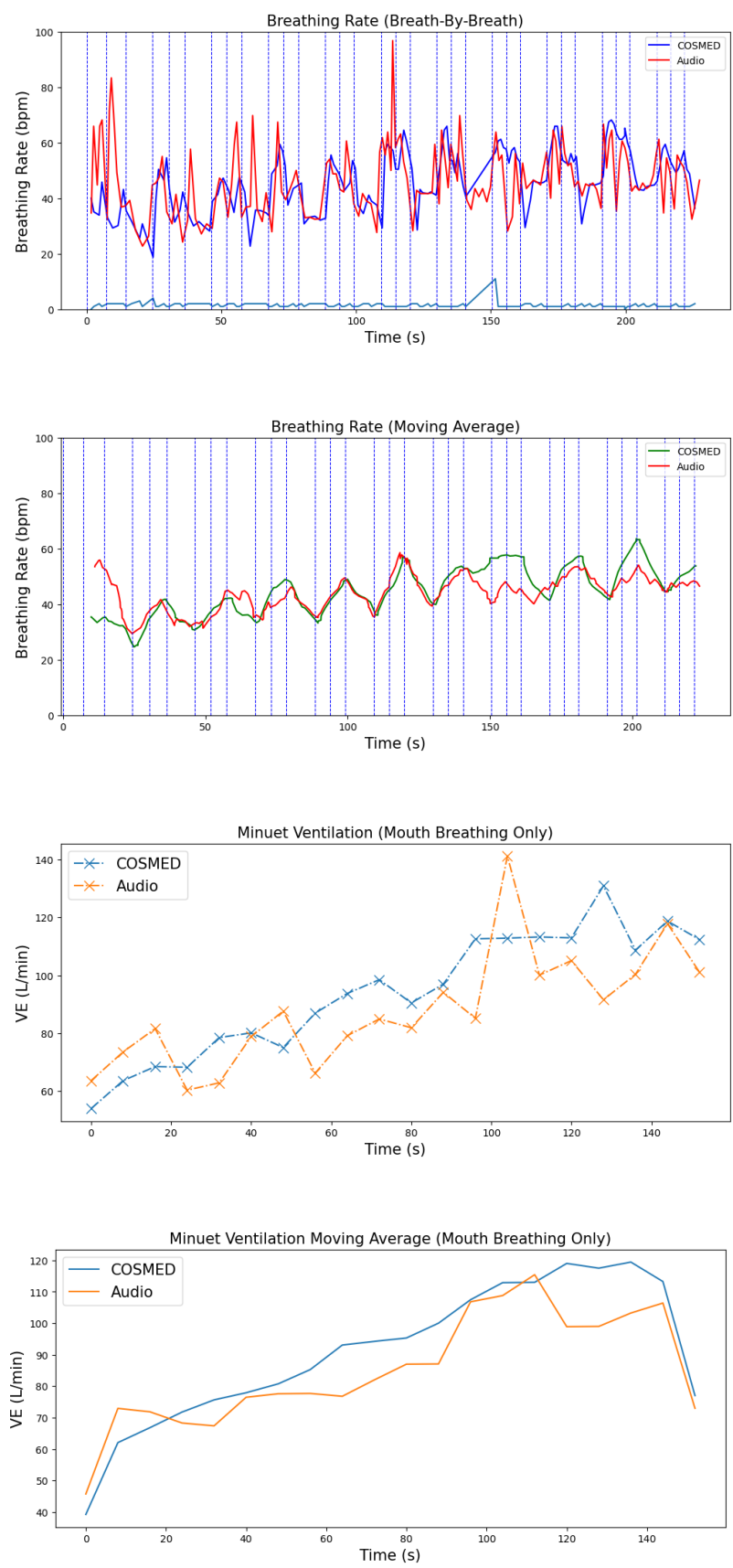


Figure D.19: Subject D14 - Test 1

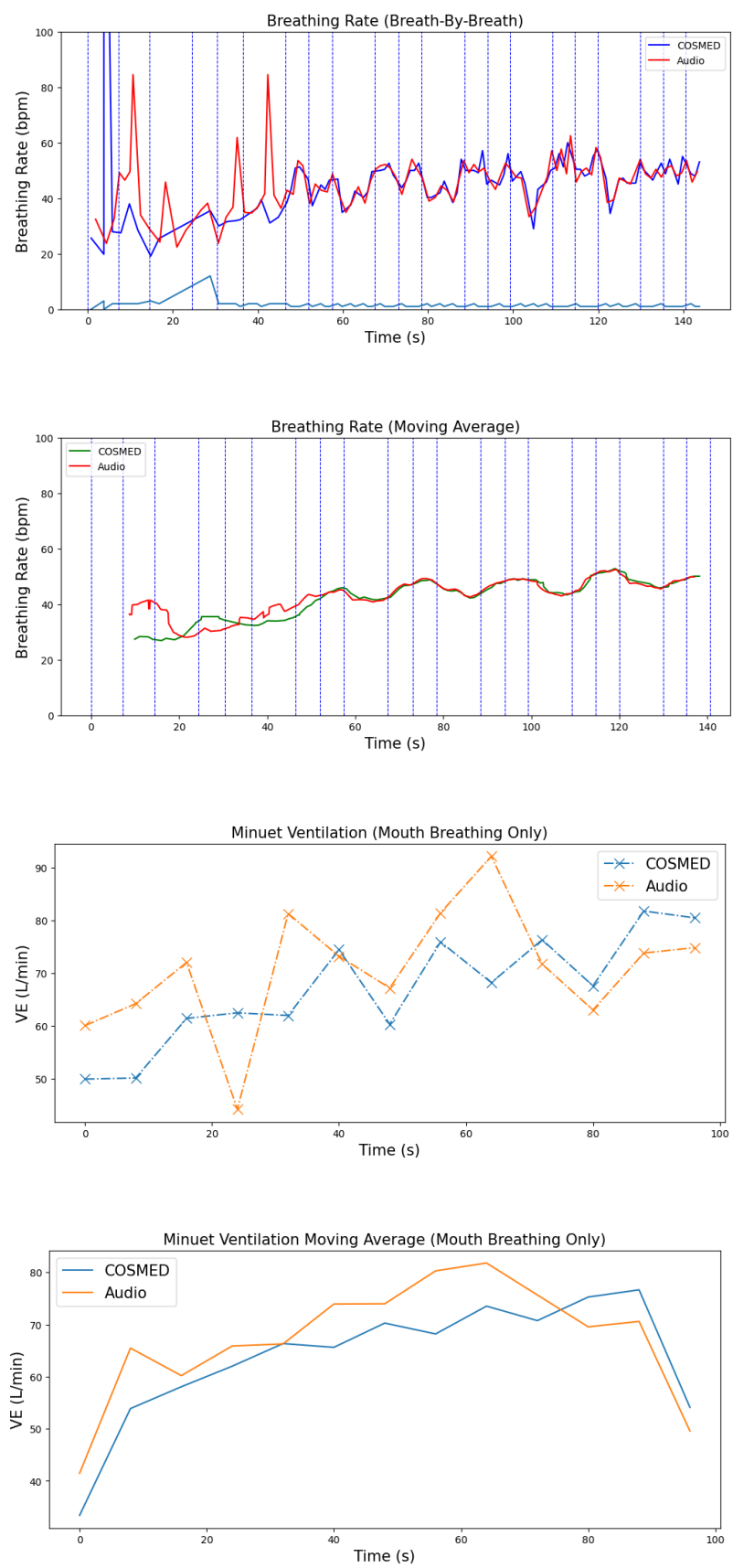


Figure D.20: Subject D14 - Test 2

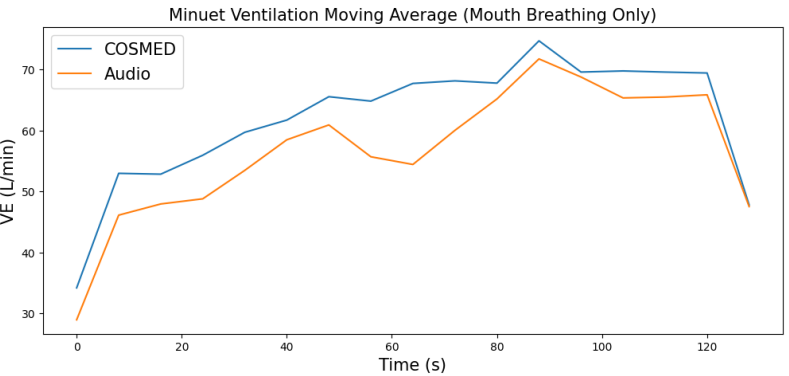
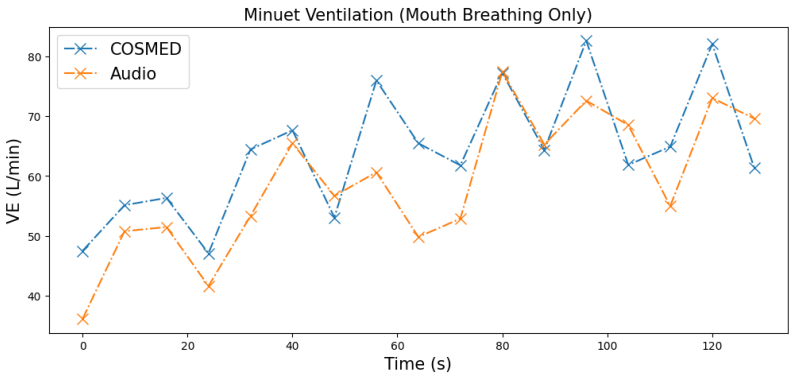
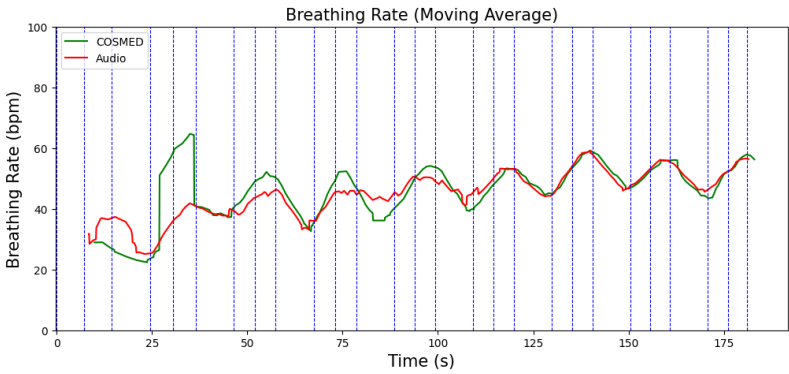
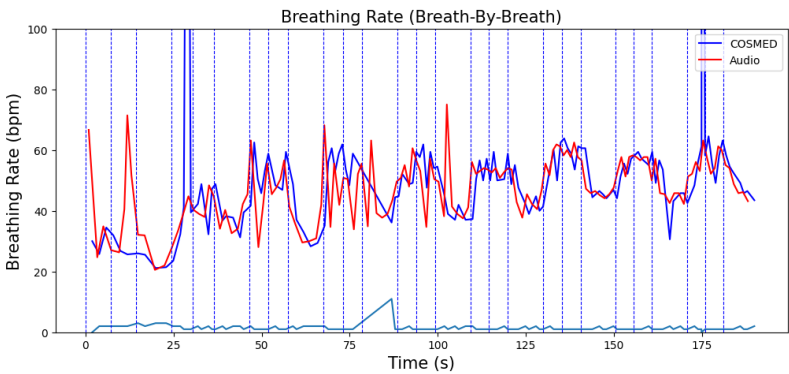


Figure D.21: Subject D15 - Test 1

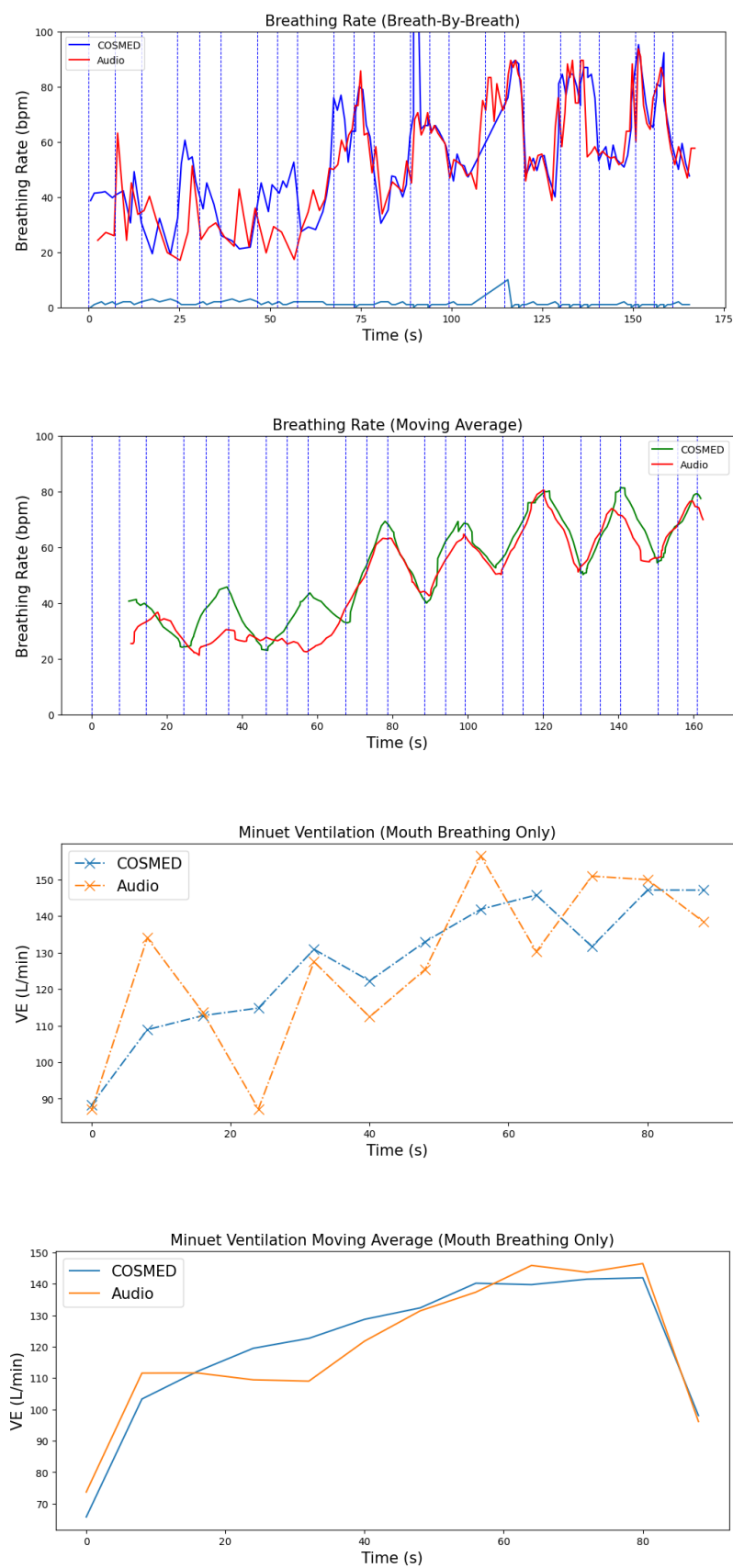


Figure D.22: Subject D15 - Test 2

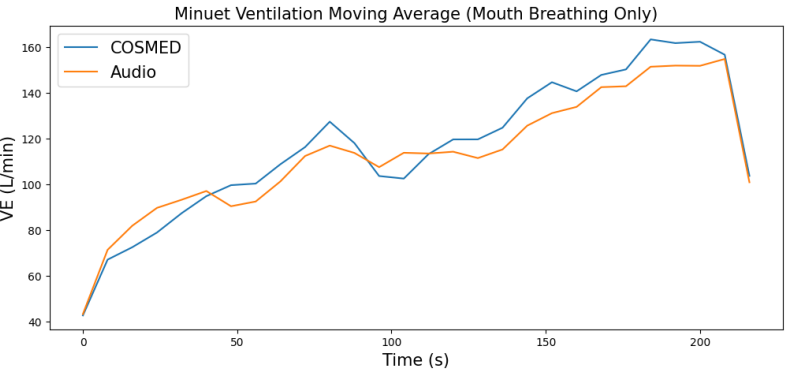
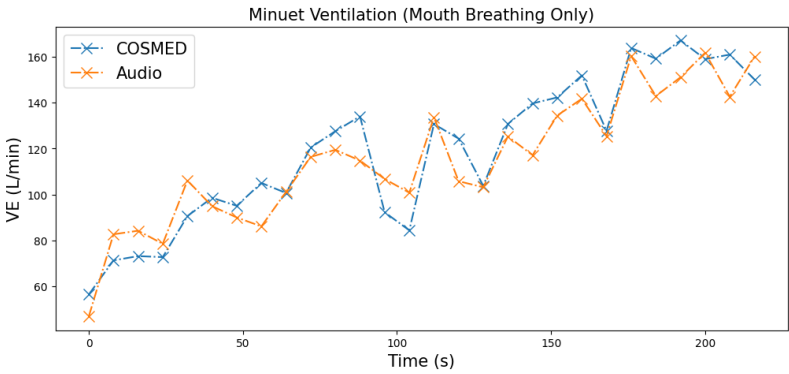
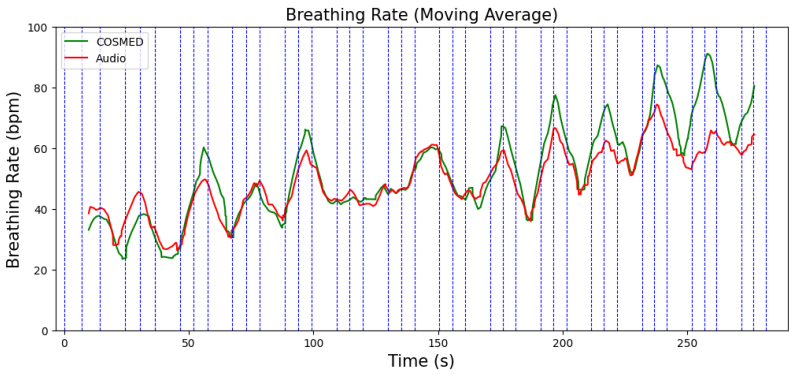
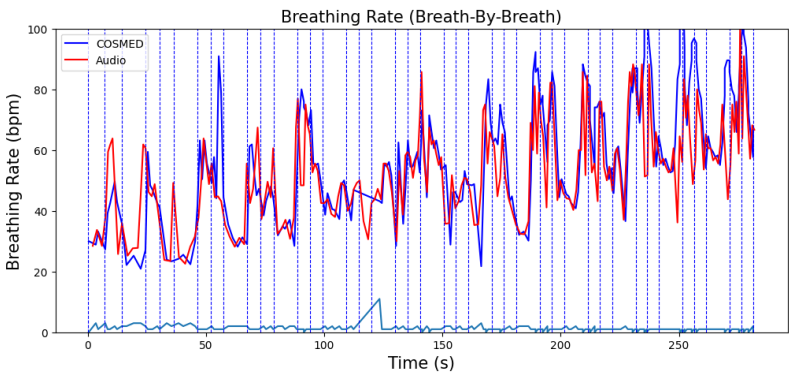


Figure D.23: Subject D16 - Test 1

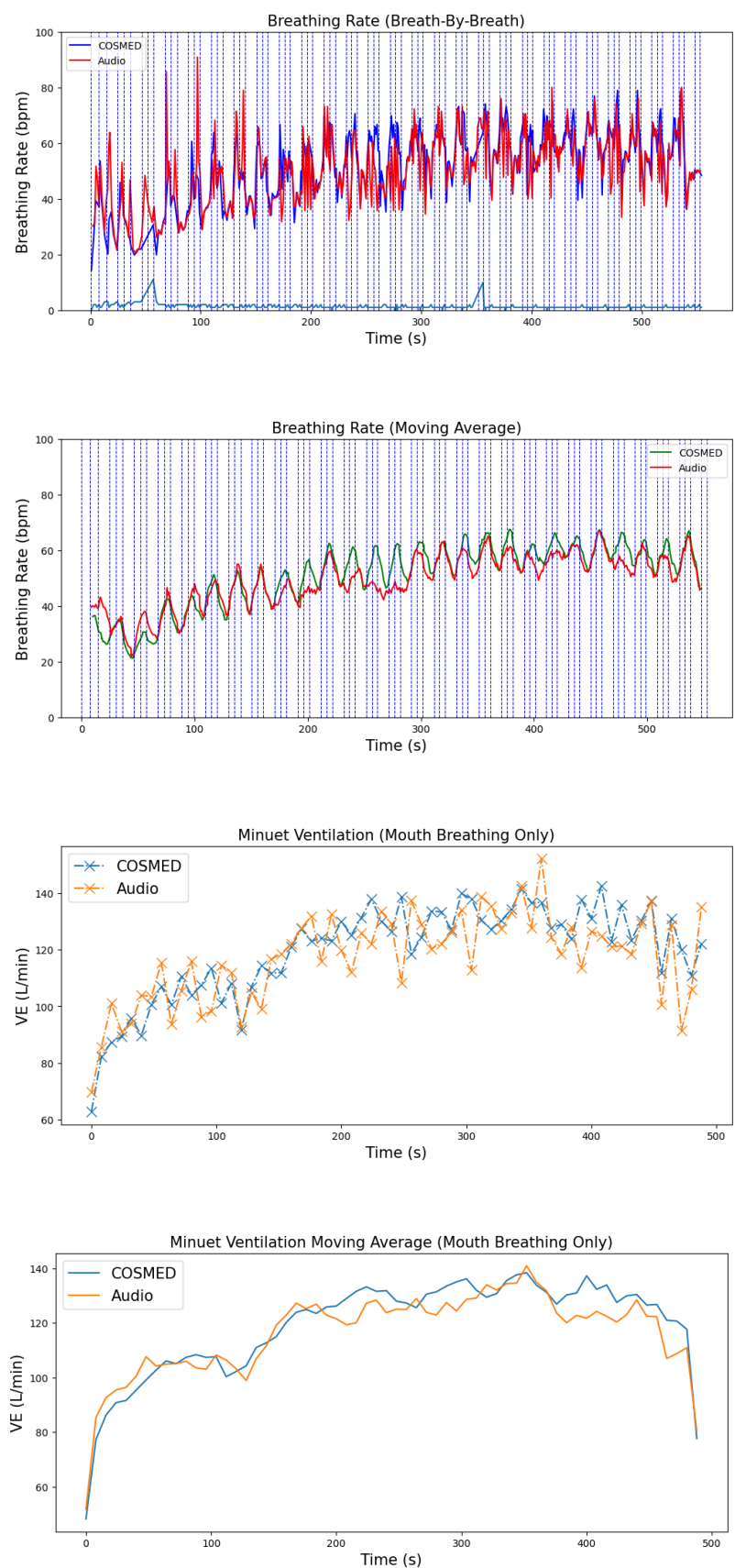


Figure D.24: Subject D16 - Test 2

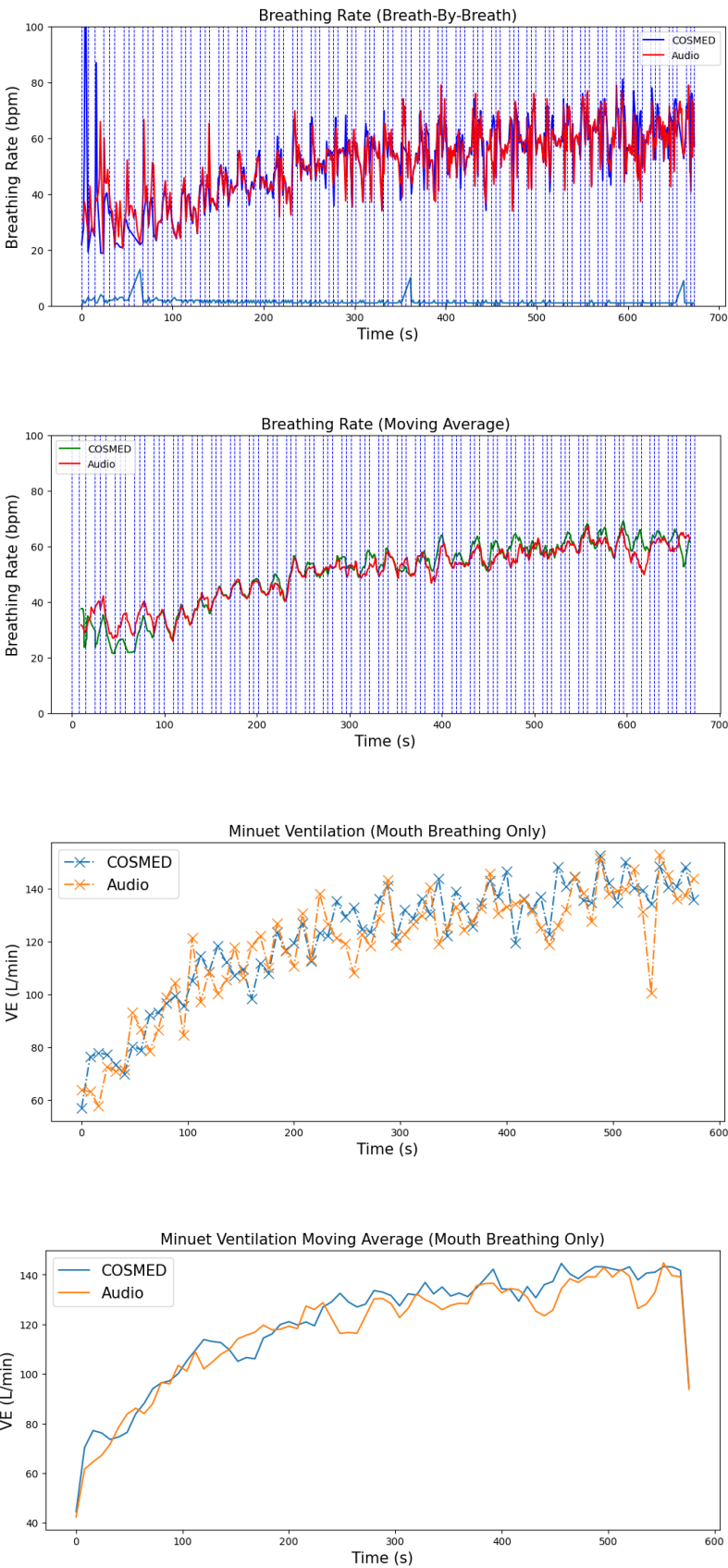


Figure D.25: Subject D19 - Test 1

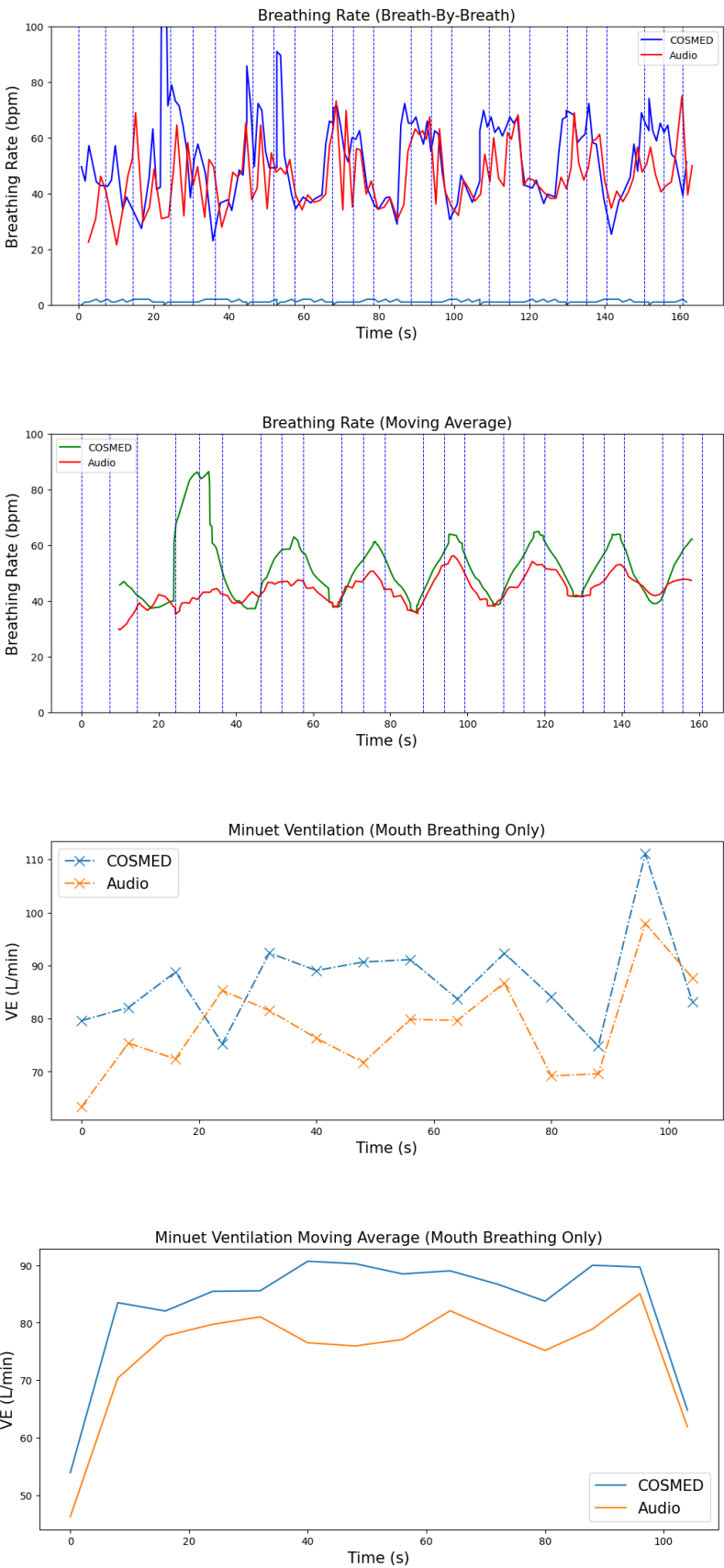


Figure D.26: Subject D20 - Test 1

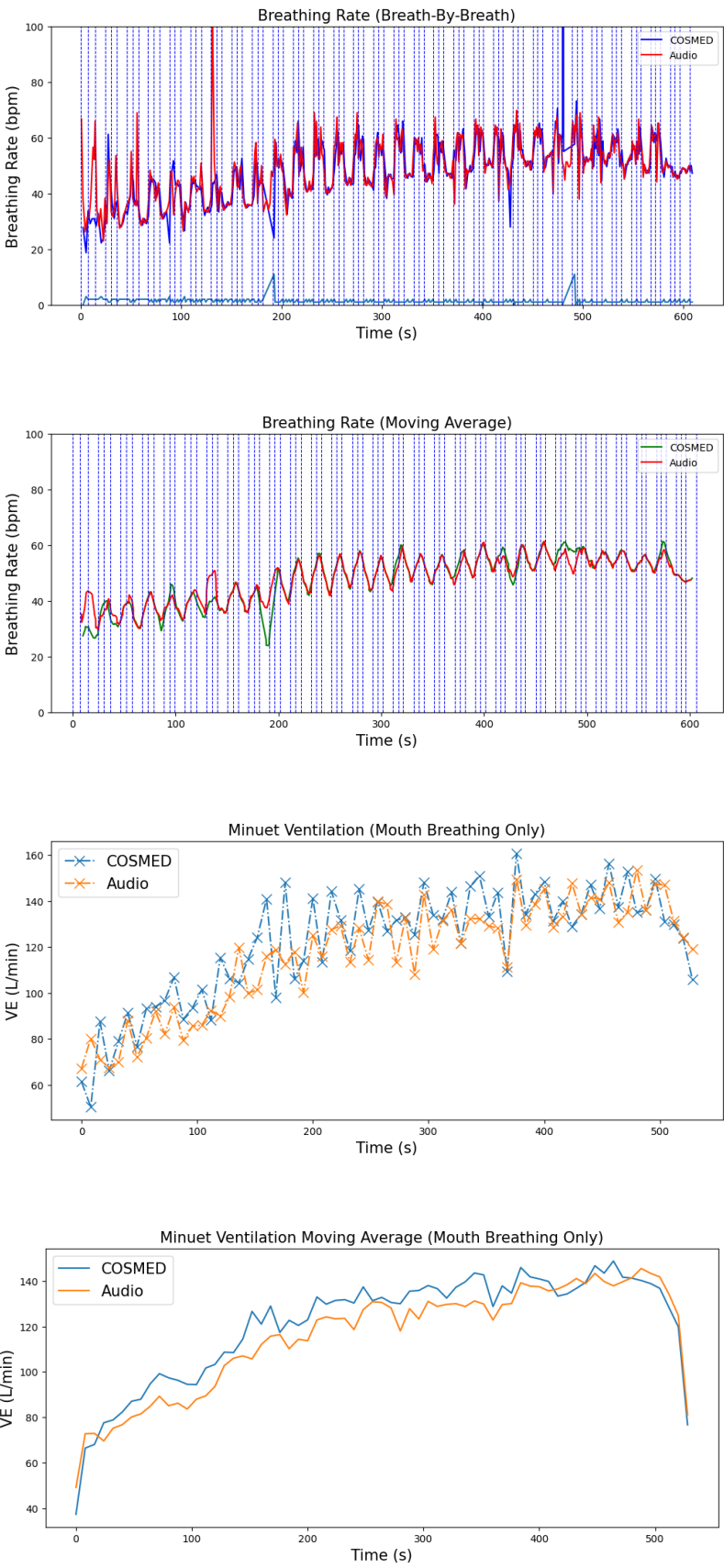


Figure D.27: Subject D20 - Test 2

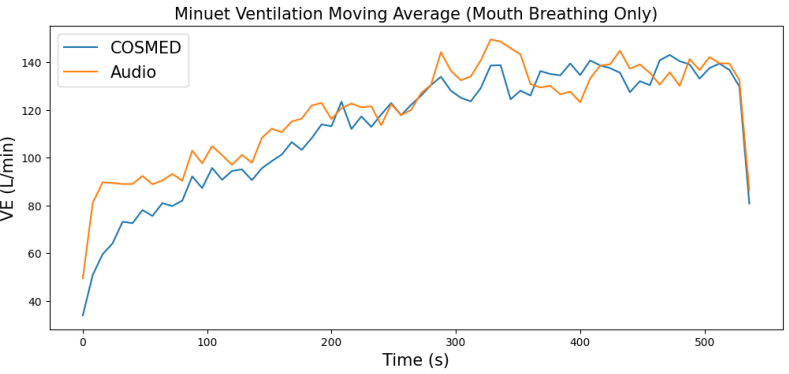
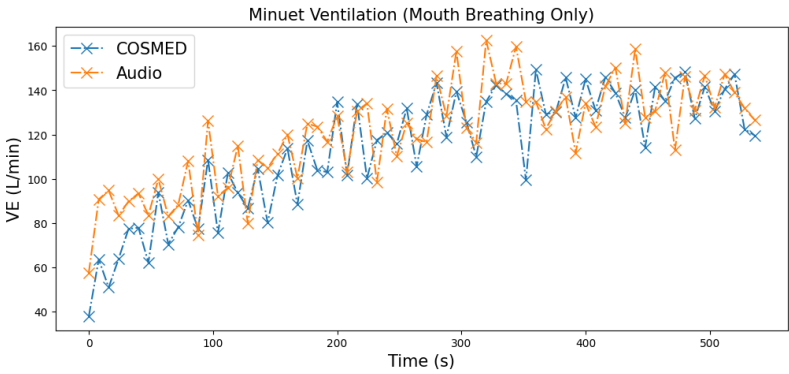
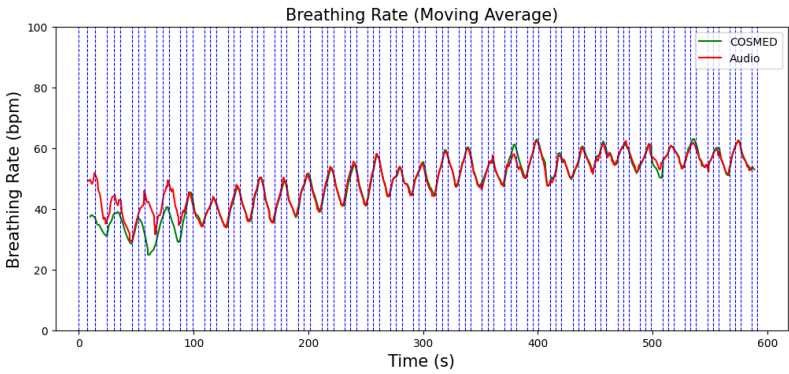
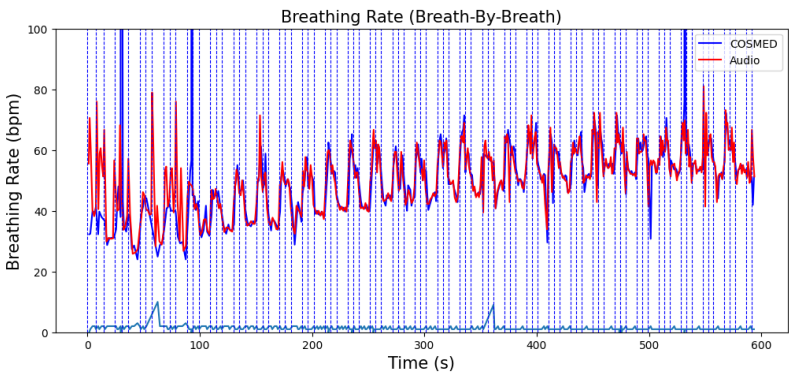


Figure D.28: Subject D21 - Test 1

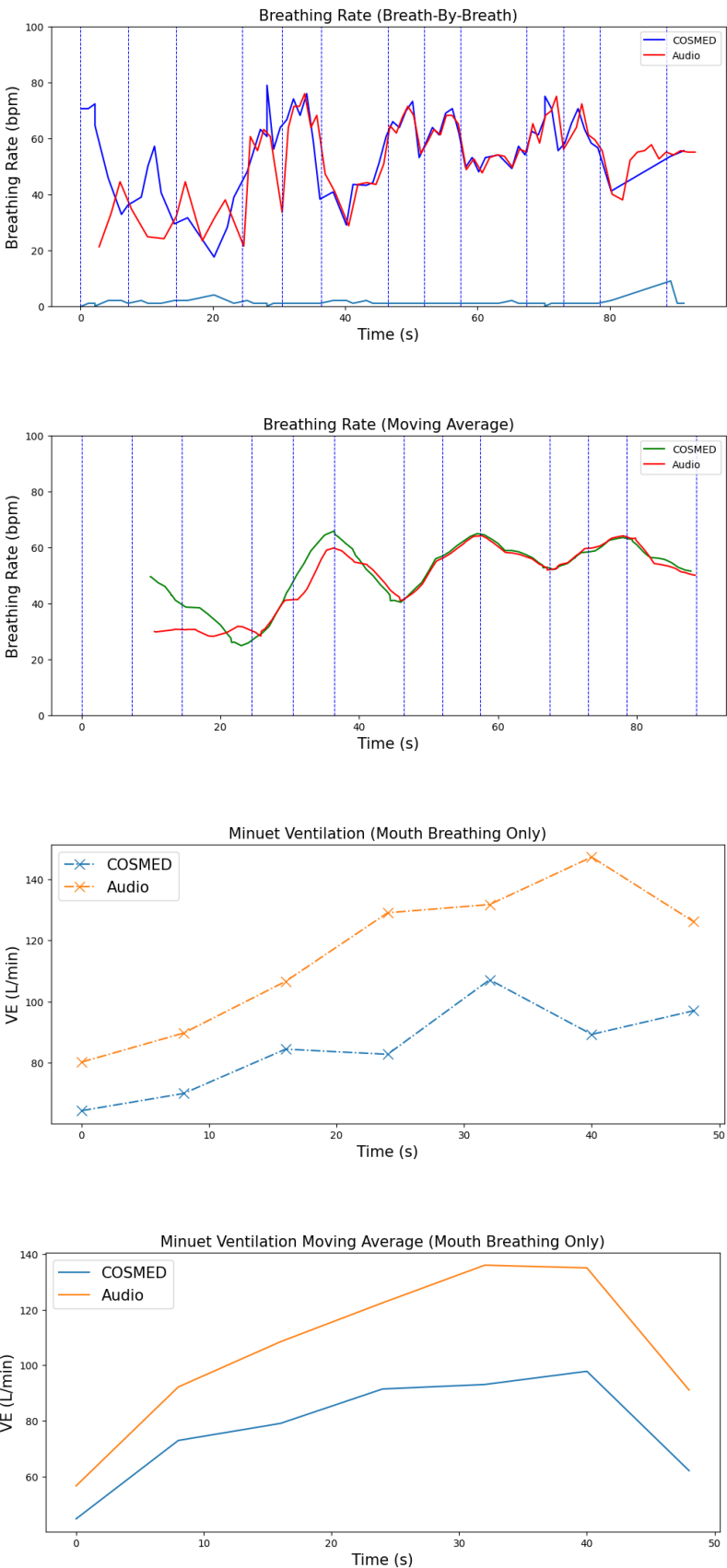


Figure D.29: Subject D24 - Test 1

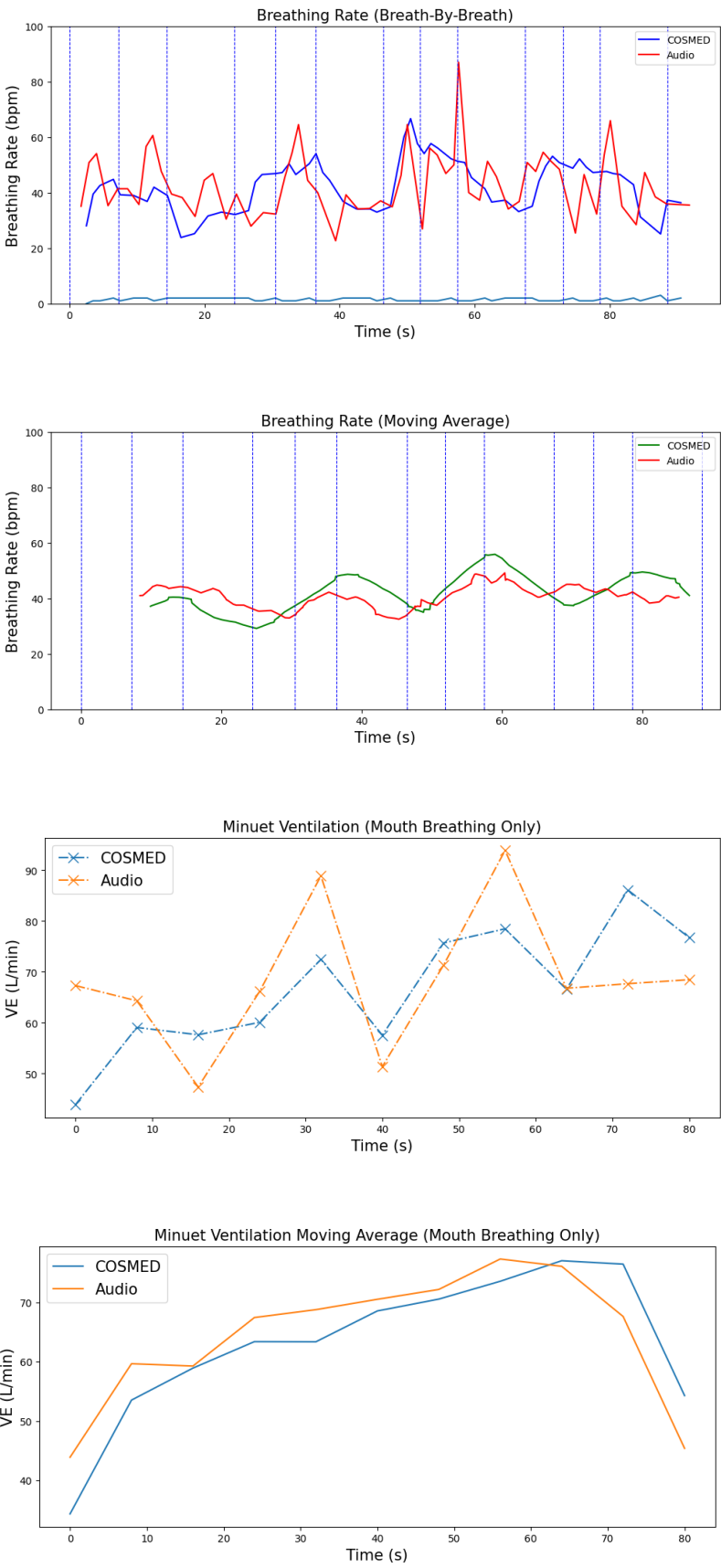
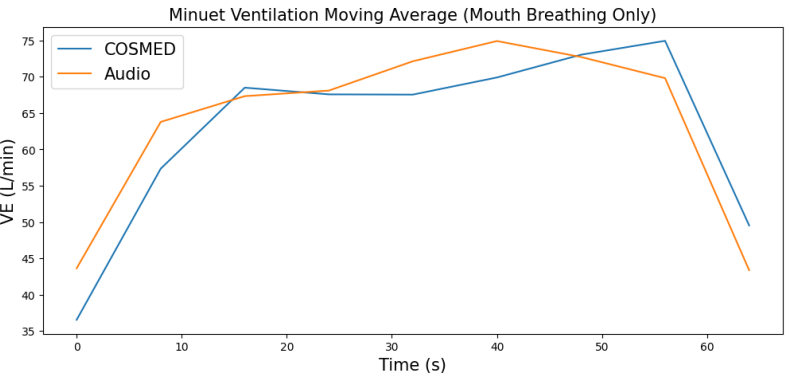
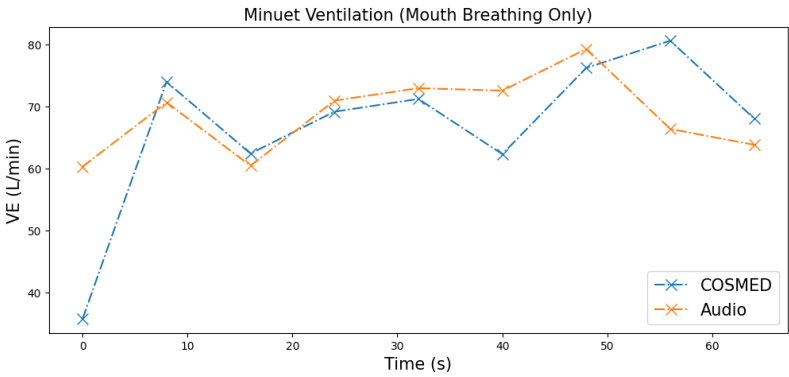
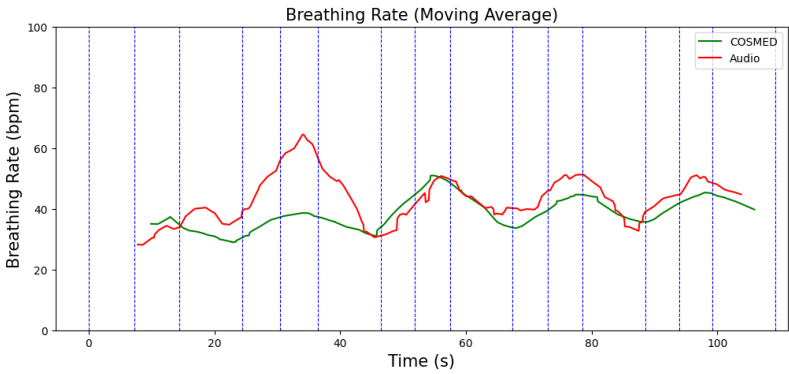
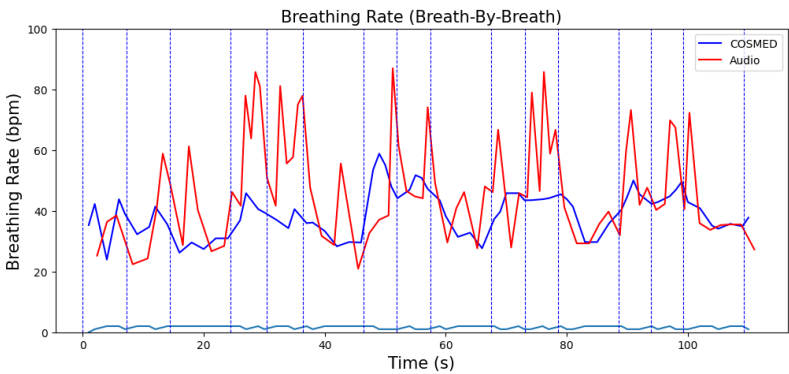


Figure D.30: Subject D24 - Test 2



Appendix E

VE Algorithms - Leave One Subject Out Results

Here we have the results for the VE algorithm leave-one-subject-out cross-validation plotted against their subject-exposed counterparts. The leave-one-subject-out cross-validation plots for Subject D11 clearly show signs of a ceiling effect. This is due to the VE range of D11 being significantly above the rest of the dataset.

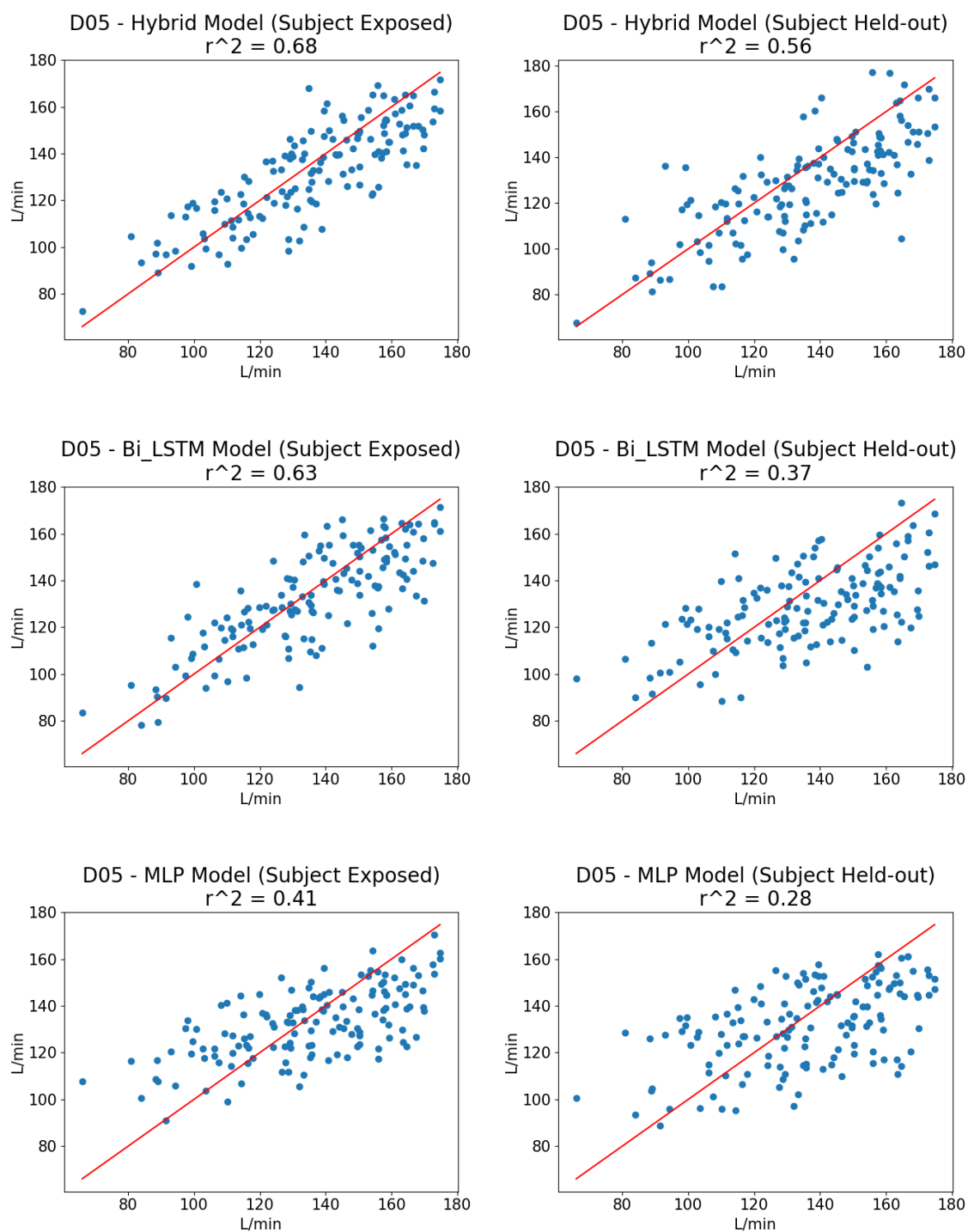


Figure E.1: D05

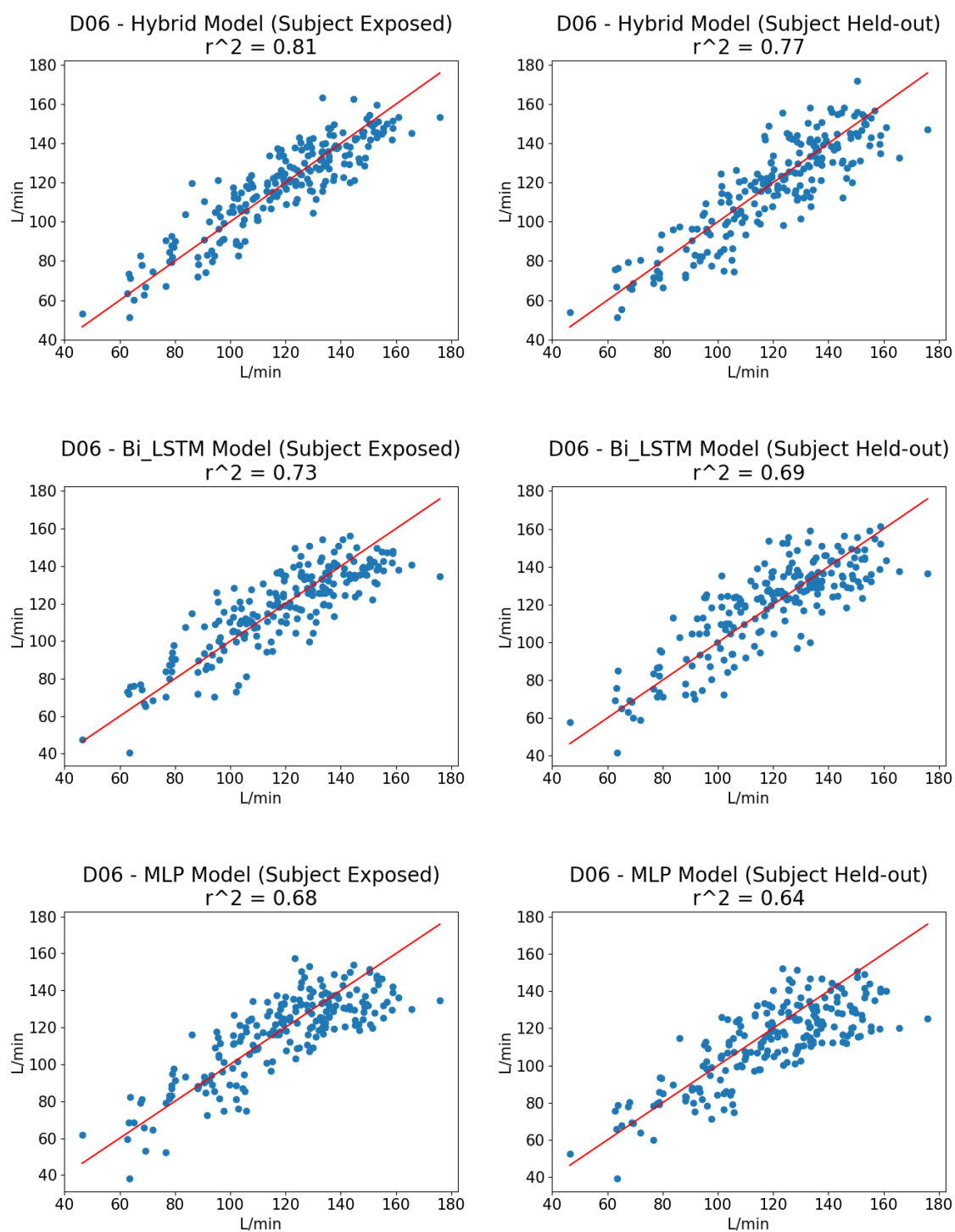


Figure E.2: D06

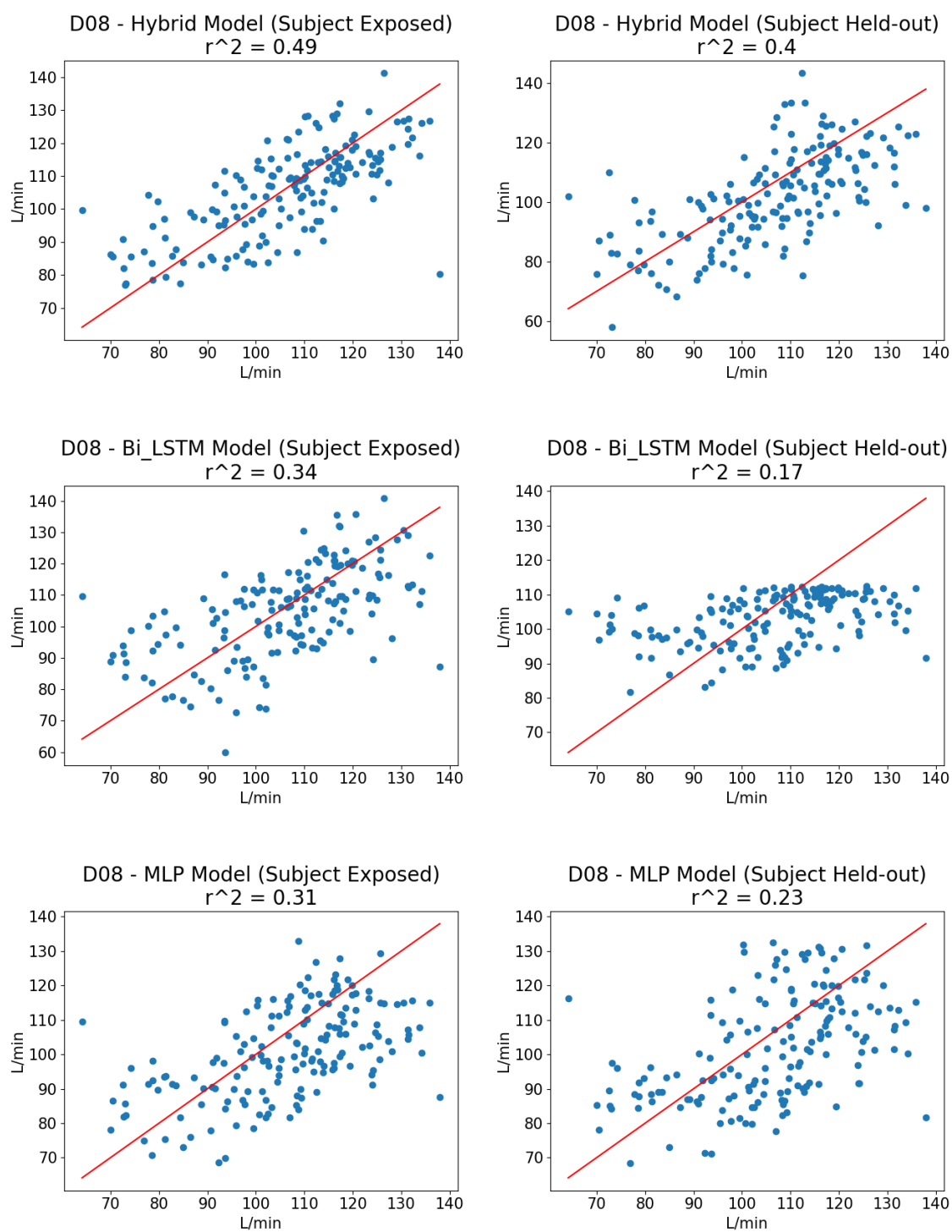


Figure E.3: D08

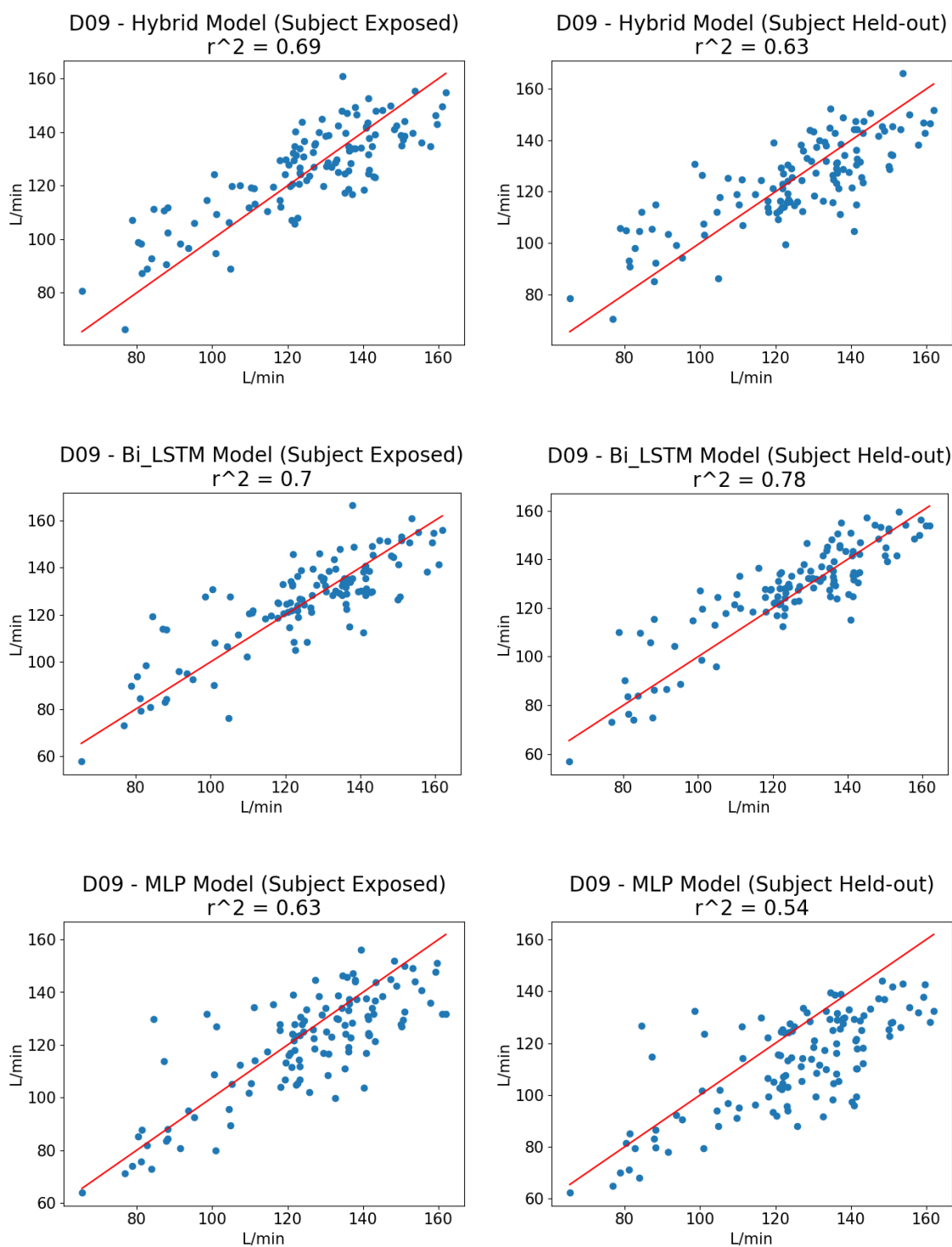


Figure E.4: D09

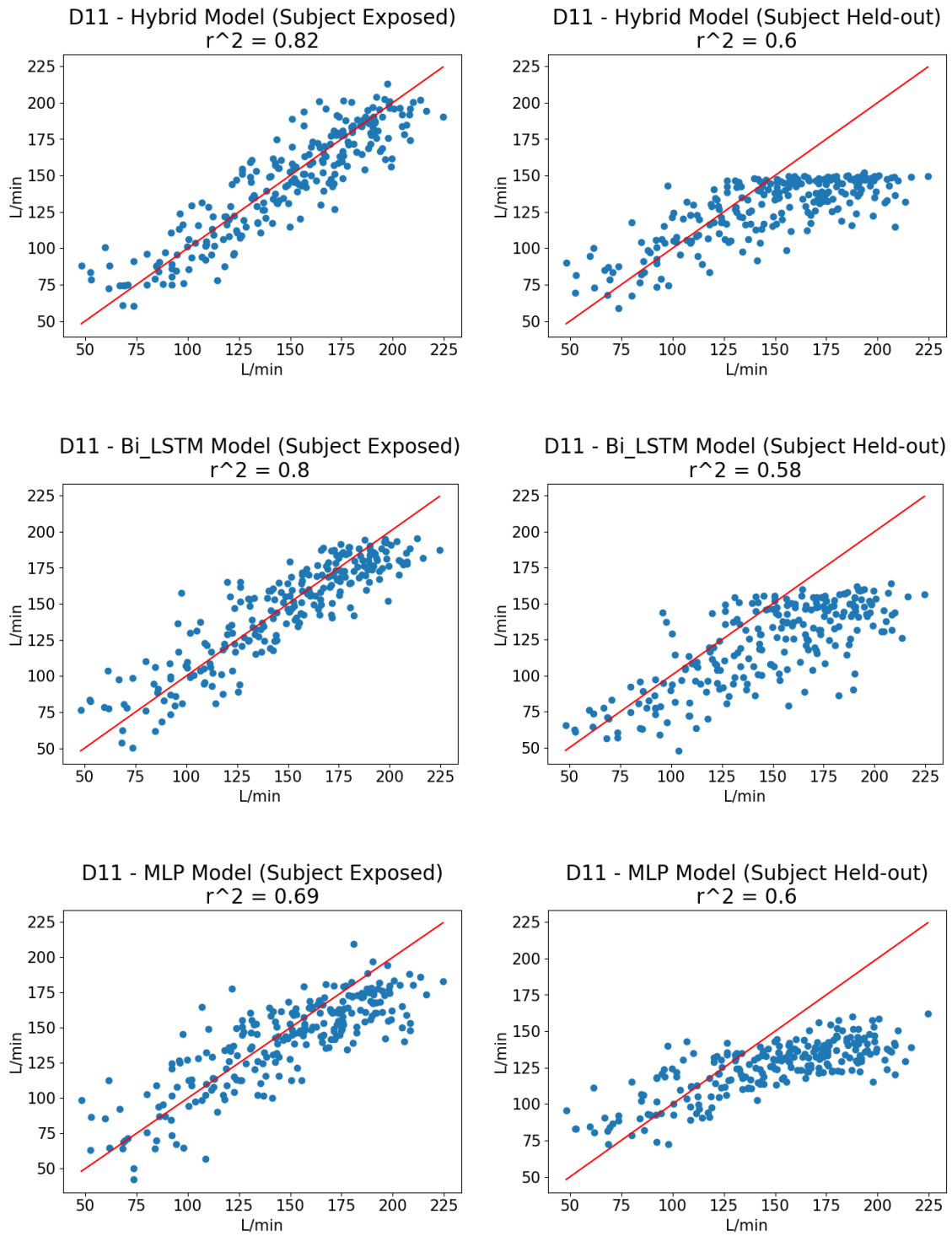


Figure E.5: D11

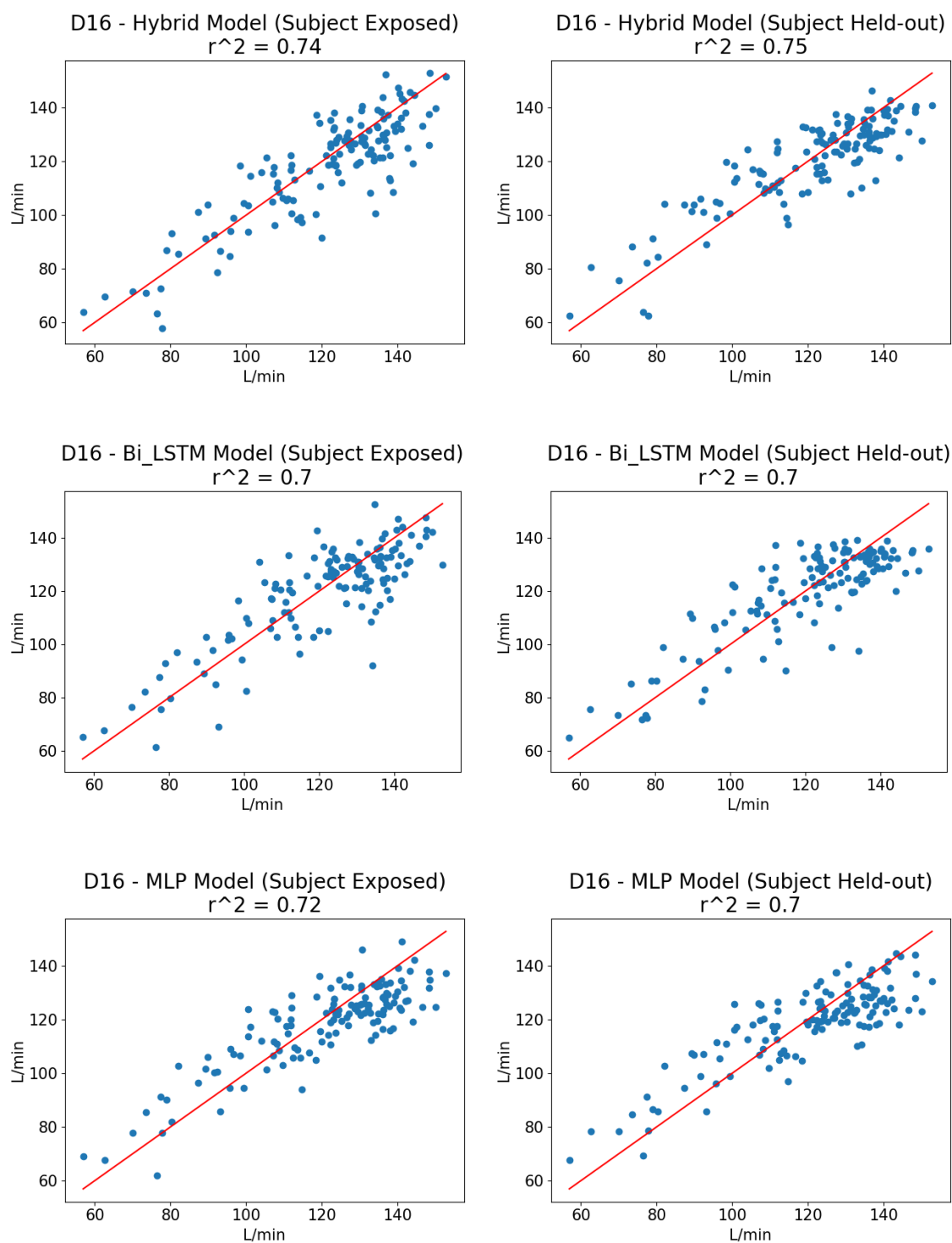


Figure E.6: D16

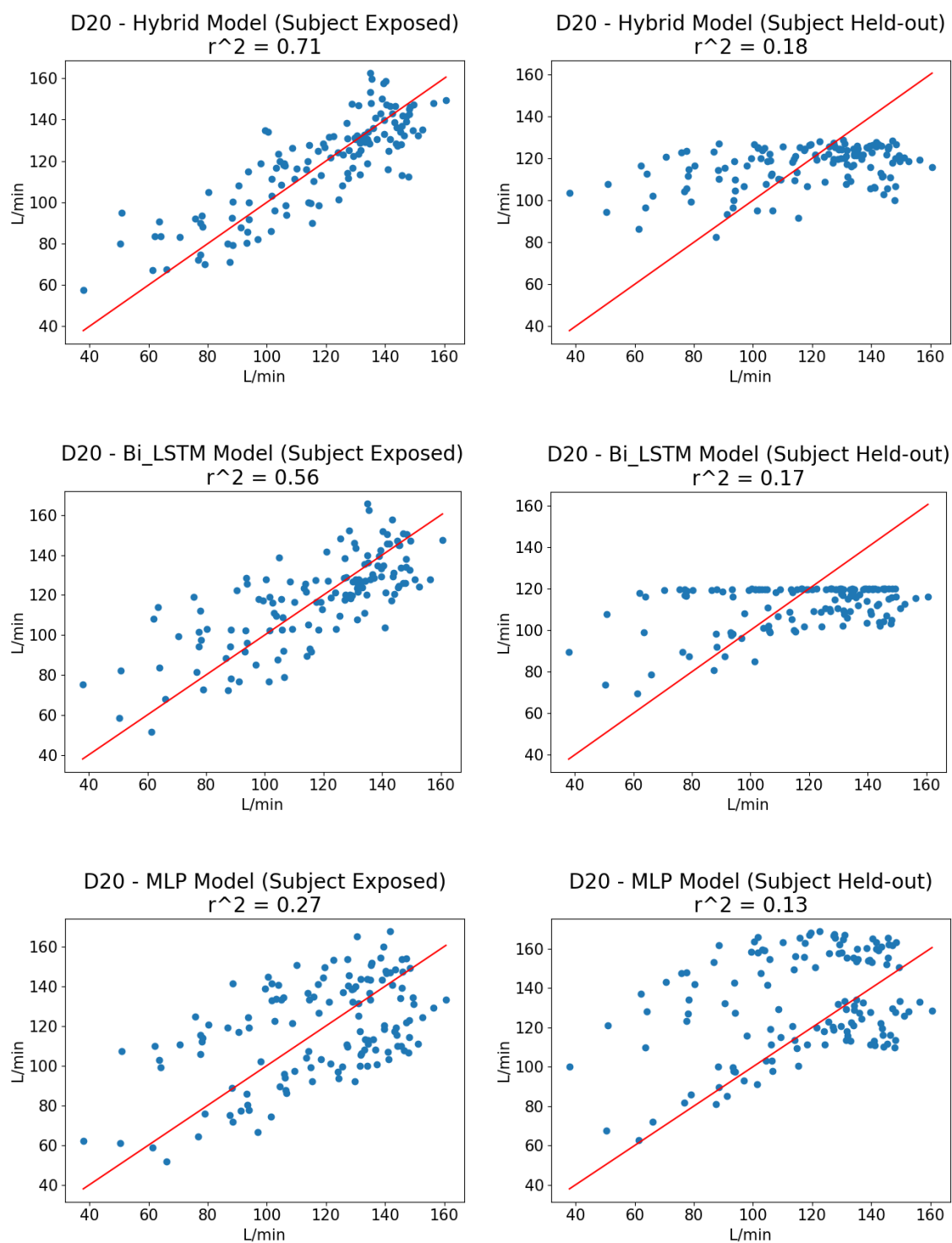


Figure E.7: D20

Appendix F

Publication Portfolio

Published Journal Articles

- Cheng R, Haste P, Levens E, Bergmann J. Feature Importance for Estimating Rating of Perceived Exertion from Cardiorespiratory Signals Using Machine Learning. *Frontiers in Sports and Active Living*. 2024 Sep 24;6:1448243.
- Cheng R, Bergmann J. A Case for an Oral Cavity-Based Respiratory Rate Sensor System. *IEEE Sensors Letters*. 2023 Mar 13;7(3):1-4.
- Cheng R, Xhameni A, Farrow T. A precision method for integrating shock sensors in the lining of sports helmets by additive manufacturing. *IEEE Sensors Letters*. 2022 Sep 8;6(10):1-4.
- Cheng R, Bergmann J. Impact and workload are dominating on-field data monitoring techniques to track health and well-being of team-sports athletes. *Physiological Measurement*. 2022 Apr 7;43(3):03TR01.
- Nagaraja VH, Cheng R, Slater DH, Thompson MS, Bergmann J. Upper-Limb Prosthetic Maintenance Data: A Retrospective Analysis Study. *JPO: Journal of Prosthetics and Orthotics*. 2022 Oct 1;34(4):223-32.

- Sharma R, Dasgupta A, Cheng R, Mishra C, Nagaraja VH. Machine Learning for Musculoskeletal Modeling of Upper Extremity. *IEEE Sensors Journal*. 2022 Aug 16;22(19):18684-97.
- de Almeida e Bueno L, Kwong MT, Milnthorpe WR, Cheng R, Bergmann J. Applying ubiquitous sensing to estimate perceived exertion based on cardiorespiratory features. *Sports Engineering*. 2021 Dec;24(1):1-9.

Pending Journal Articles

- Cheng R, Bergmann J. Estimating Breathing Rate and Minute Ventilation During Running Using Breathing Audio Obtained Through A Microphone-Instrumented Mouthguard. Submitted to *Wearable Technologies*.
- Cheng R, Nagaraja VH, Berthaume MA, Dickinson A, Oldfrey B, Dasgupta, A, Kenney LPJ. Co-creating a Schema for a Prosthetic-limb Provision and Repair Registry: A Data Standardization Proposal. Submitted to *Archives of Rehabilitation Research and Clinical Translation*.
- Nagaraja VH, Kumar PA, Moulic SG, Cheng R, Varadhan M, Bergmann J. A Nearly Two-decade-long Maintenance and Cost Analysis of Prosthetic arms in a Low-resource Setting: Case for More Appropriate Medical Device Service Systems.
- Villar, DA, Cheng R, Andrew G. Niche Change, Climatic Niche, Holocene, Pleistocene, Bird, Dispersal Syndrome. Submitted to *Palaeontology*

Conferences & Workshops

- Nagaraja, VH, Cheng, R, and Kenney, L. Prosthetic-limb Provision and Repair Registry Schema. British Association Of Prosthetists & Orthotists (BAPO), Wigan, Greater Manchester, UK. 16th May, 2025

- Nagaraja, VH, Cheng, R, and Kenney, L. Co-creation of an AT Provision and Repair Registry: A Step Towards Data-driven Rehabilitation Service Provision. CREATE Conference on Complex Rehab Technology. Mobility India, Bangalore, India. 18th - 19th October, 2024
- Nagaraja, VH, Cheng, R, and Kenney, L. Prosthetic Provision and Repair Registry – Consensus Survey. TIDAL P&O Repairs Workshop. University of Salford, Salford Media City, UK. 16th May 2024

Conferences Attended Prior to The DPhil

- Cheng R, Nagaraja VH, Bergmann J, Thompson MS. An Opensource Framework to Analyse Marker-based and Inertial-Sensor-based Measurements: Motion Capture Analysis & Plotting Assistant (MCAPA) 2.0. BioMedEng'19 Conference. Imperial College London, UK. 5th – 6th September 2019
- Cheng R, Nagaraja VH, Bergmann J, Thompson MS. Motion Capture Analysis & Plotting Assistant: An Opensource Framework to Analyse Inertial-Sensor-based Measurements. Trent International Prosthetics Symposium - 2019 (TIPS'19), Manchester, UK. 20th – 23rd March 2019
- Nagaraja VH, Cheng R, Kwong E, Bergmann J, Andersen MS, Thompson MS. Marker-based vs Inertial-based Motion Capture: Musculoskeletal Modelling of Upper Extremity Kinetics. Trent International Prosthetics Symposium - 2019 (TIPS'19), Manchester, UK. 20th – 23rd March 2019

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. & Zheng, X. (2015), ‘TensorFlow: Large-scale machine learning on heterogeneous systems’. Software available from tensorflow.org.

URL: <http://tensorflow.org/>

Abbiss, C. R. & Laursen, P. B. (2005), ‘Models to explain fatigue during prolonged endurance cycling’, *Sports medicine* **35**, 865–898.

Ackerman, I. N., Bohensky, M. A., Kemp, J. L. & de Steiger, R. (2019), ‘Likelihood of knee replacement surgery up to 15 years after sports injury: a population-level data linkage study’, *Journal of science and medicine in sport* **22**(6), 629–634.

Alam, R., Peden, D. B. & Lach, J. C. (2020), ‘Wearable respiration monitoring: interpretable inference with context and sensor biomarkers’, *IEEE journal of biomedical and health informatics* **25**(6), 1938–1948.

Albert, J. A., Herdick, A., Brahms, C. M., Granacher, U. & Arnrich, B. (2021), Using machine learning to predict perceived exertion during resistance training with

- wearable heart rate and movement sensors, *in* ‘2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)’, IEEE, pp. 801–808.
- Aliverti, A. (2017), ‘Wearable technology: role in respiratory health and disease’, *Breathe* **13**(2), e27–e36.
- Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Cheng, Q., Chen, G. et al. (2016), Deep speech 2: End-to-end speech recognition in english and mandarin, *in* ‘International conference on machine learning’, PMLR, pp. 173–182.
- Antonelli, A., Guilizzoni, D., Angelucci, A., Melloni, G., Mazza, F., Stanzi, A., Venturino, M., Kuller, D. & Aliverti, A. (2020), ‘Comparison between the airgo™ device and a metabolic cart during rest and exercise’, *Sensors* **20**(14), 3943.
- Apostolou, K. & Tjortjis, C. (2019), Sports analytics algorithms for performance prediction, *in* ‘2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)’, IEEE, pp. 1–4.
- Atkins, S. J. (2006), ‘Performance of the yo-yo intermittent recovery test by elite professional and semiprofessional rugby league players’, *The Journal of Strength & Conditioning Research* **20**(1), 222–225.
- Axelrod, D., Ziegler, T., Pincus, D., Widdifield, J., Marks, P., Paterson, M. & Wasserstein, D. (2022), ‘Basketball vs. hockey—the changing face of sport-related injuries in canada’, *Clinical journal of sport medicine* **32**(3), e281–e287.
- Bailes, J. E. & Cantu, R. C. (2001), ‘Head injury in athletes’, *Neurosurgery* **48**(1), 26–46.
- Bailes, J. E., Petraglia, A. L., Omalu, B. I., Nauman, E. & Talavage, T. (2013), ‘Role

- of subconcussion in repetitive mild traumatic brain injury: a review', *Journal of neurosurgery* **119**(5), 1235–1245.
- Bailey, A., Funk, J., Lessley, D., Sherwood, C., Crandall, J., Neale, W. & Rose, N. (2020), 'Validation of a videogrammetry technique for analysing american football helmet kinematics', *Sports biomechanics* **19**(5), 678–700.
- Bangsbo, J., Iaia, F. M. & Krstrup, P. (2008), 'The yo-yo intermittent recovery test: a useful tool for evaluation of physical performance in intermittent sports', *Sports medicine* **38**, 37–51.
- Bari, S., Svaldi, D. O., Jang, I., Shenk, T. E., Poole, V. N., Lee, T., Dydak, U., Rispoli, J. V., Nauman, E. A. & Talavage, T. M. (2019), 'Dependence on subconcussive impacts of brain metabolism in collision sport athletes: an mr spectroscopic study', *Brain Imaging and Behavior* **13**, 735–749.
- Bartsch, A., Dama, R., Alberts, J., Samorezov, S., Benzel, E., Miele, V., Shah, A., Humm, J., McCrea, M. & Stemper, B. (2020), 'Measuring blunt force head impacts in athletes', *Military medicine* **185**(Supplement_1), 190–196.
- Benchetrit, G. (2000), 'Breathing pattern in humans: diversity and individuality', *Respiration physiology* **122**(2-3), 123–129.
- Benson, L. C., Räisänen, A. M., Volkova, V. G., Pasanen, K. & Emery, C. A. (2020), 'Workload a-wear-ness: Monitoring workload in team...', *Journal of Orthopaedic and Sports Physical Therapy* **50**(10), 549–563.
- Bernardi, L., Spadacini, G., Bellwon, J., Hajric, R., Roskamm, H. & Frey, A. W. (1998), 'Effect of breathing rate on oxygen saturation and exercise performance in chronic heart failure', *The Lancet* **351**(9112), 1308–1311.

- Boersma, P. (2001), ‘Praat, a system for doing phonetics by computer’, *Glott. Int.* **5**(9), 341–345.
- Boiten, F. (1993), ‘Component analysis of task-related respiratory patterns’, *International Journal of psychophysiology* **15**(2), 91–104.
- Borg, G. (1990), ‘Psychophysical scaling with applications in physical work and the perception of exertion’, *Scandinavian journal of work, environment & health* pp. 55–58.
- Borg, G. (1998), *Borg’s perceived exertion and pain scales.*, Human kinetics.
- Boulding, R., Stacey, R., Niven, R. & Fowler, S. J. (2016), ‘Dysfunctional breathing: a review of the literature and proposal for classification’, *European Respiratory Review* **25**(141), 287–294.
- Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., Gabbett, T. J., Coutts, A. J., Burgess, D. J., Gregson, W. et al. (2017), ‘Monitoring athlete training loads: consensus statement’, *International journal of sports physiology and performance* **12**(s2), S2–161.
- Boyd, L. J., Ball, K. & Aughey, R. J. (2011), ‘The reliability of minimaxx accelerometers for measuring physical activity in australian football’, *International journal of sports physiology and performance* **6**(3), 311–321.
- Breiman, L. (2001), ‘Random forests’, *Machine learning* **45**, 5–32.
- Brennan, J. H., Mitra, B., Synnot, A., McKenzie, J., Willmott, C., McIntosh, A. S., Maller, J. J. & Rosenfeld, J. V. (2017), ‘Accelerometers for the assessment of concussion in male athletes: a systematic review and meta-analysis’, *Sports medicine* **47**, 469–478.

- Bridgman, H., Kwong, M. T. & Bergmann, J. H. (2019), ‘Mechanical safety of embedded electronics for in-body wearables: A smart mouthguard study’, *Annals of biomedical engineering* **47**(8), 1725–1737.
- Buice, J. M., Esquivel, A. O. & Andreovich, C. J. (2018), ‘Laboratory validation of a wearable sensor for the measurement of head acceleration in men’s and women’s lacrosse’, *Journal of biomechanical engineering* **140**(10), 101004.
- Buttar, K. K., Saboo, N. & Kacker, S. (2019), ‘A review: Maximal oxygen uptake (vo2 max) and its estimation methods’, *Ijpes* **6**(6), 24–32.
- Caccese, J. B., Best, C., Lamond, L. C., DiFabio, M., Kaminski, T. W., Watson, D., Getchell, N. & Buckley, T. A. (2019), ‘Effects of repetitive head impacts on a concussion assessment battery.’, *Medicine and science in sports and exercise* **51**(7), 1355–1361.
- Cameron, M. & MacDougall, C. (2000), ‘Crime prevention through sport and physical activity.’, *Trends & Issues in Crime & Criminal Justice* (165).
- Campbell, K. R., Warnica, M. J., Levine, I. C., Brooks, J. S., Laing, A. C., Burkhart, T. A. & Dickey, J. P. (2016), ‘Laboratory evaluation of the gforce tracker™, a head impact kinematic measuring device for use in football helmets’, *Annals of biomedical engineering* **44**, 1246–1256.
- Cardinale, M. & Varley, M. C. (2017), ‘Wearable training-monitoring technology: applications, challenges, and opportunities’, *International journal of sports physiology and performance* **12**(s2), S2–55.
- Cardoos, N. (2015), ‘Overtraining syndrome’, *Current sports medicine reports* **14**(3), 157–158.

- Carey, D. L., Ong, K., Morris, M. E., Crow, J. & Crossley, K. M. (2016), ‘Predicting ratings of perceived exertion in australian football players: methods for live estimation’, *International Journal of Computer Science in Sport* **15**(2), 64–77.
- Castagna, C., Impellizzeri, F. M., Chamari, K., Carlomagno, D. & Rampinini, E. (2006), ‘Aerobic fitness and yo-yo continuous and intermittent tests performances in soccer players: a correlation study’, *The Journal of Strength & Conditioning Research* **20**(2), 320–325.
- Castagna, C., Impellizzeri, F. M., Rampinini, E., D’Ottavio, S. & Manzi, V. (2008), ‘The yo-yo intermittent recovery test in basketball players’, *Journal of science and medicine in sport* **11**(2), 202–208.
- Castellano, J., Alvarez-Pastor, D. & Bradley, P. S. (2014), ‘Evaluation of research using computerised tracking systems (amisco® and prozone®) to analyse physical performance in elite soccer: A systematic review’, *Sports medicine* **44**, 701–712.
- Chang, H.-H., Hsu, C.-C., Chen, C.-Y., Lee, W.-K., Hsu, H.-T., Shyu, K.-K., Yeh, J.-R., Lin, P.-J. & Lee, P.-L. (2018), ‘A method for respiration rate detection in wrist ppg signal using holo-hilbert spectrum’, *IEEE Sensors Journal* **18**(18), 7560–7569.
- Charlton, P. H., Birrenkott, D. A., Bonnici, T., Pimentel, M. A., Johnson, A. E., Alastruey, J., Tarassenko, L., Watkinson, P. J., Beale, R. & Clifton, D. A. (2017), ‘Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review’, *IEEE reviews in biomedical engineering* **11**, 2–20.
- Cheng, R. & Bergmann, J. (2023), ‘A case for an oral cavity-based respiratory rate sensor system’, *IEEE Sensors Letters* **7**(3), 1–4.
- Cheng, R. & Bergmann, J. H. (2022), ‘Impact and workload are dominating on-field data monitoring techniques to track health and well-being of team-sports athletes’, *Physiological Measurement* **43**(3), 03TR01.

- Cheng, R., Haste, P., Levens, E. & Bergmann, J. (2024), ‘Feature importance for estimating rating of perceived exertion from cardiorespiratory signals using machine learning’, *Frontiers in Sports and Active Living* **6**, 1448243.
- Chowdhury, A. K., Tjondronegoro, D., Chandran, V., Zhang, J. & Trost, S. G. (2019), ‘Prediction of relative physical activity intensity using multimodal sensing of physiological data’, *Sensors* **19**(20), 4509.
- Chu, M., Nguyen, T., Pandey, V., Zhou, Y., Pham, H. N., Bar-Yoseph, R., Radom-Aizik, S., Jain, R., Cooper, D. M. & Khine, M. (2019), ‘Respiration rate and volume measurements using wearable strain sensors’, *NPJ digital medicine* **2**(1), 8.
- Clemente, F. M., Mendes, B., Palao, J. M., Silvério, A., Carriço, S., Calvete, F. & Nakamura, F. Y. (2018), ‘Seasonal player wellness and its longitudinal association with internal training load: study in elite volleyball.’, *The Journal of sports medicine and physical fitness* **59**(3), 345–351.
- Clifford, G. D., Azuaje, F., McSharry, P. et al. (2006), *Advanced methods and tools for ECG data analysis*, Vol. 10, Artech house Boston.
- Colosio, A. L., Lievens, M., Pogliaghi, S., Bourgois, J. G. & Boone, J. (2020), ‘Heart rate-index estimates aerobic metabolism in professional soccer players’, *Journal of science and medicine in sport* **23**(12), 1208–1214.
- Costanzo, I., Sen, D., Rhein, L. & Guler, U. (2020), ‘Respiratory monitoring: Current state of the art and future roads’, *IEEE Reviews in Biomedical Engineering* **15**, 103–121.
- Cretikos, M. A., Bellomo, R., Hillman, K., Chen, J., Finfer, S. & Flabouris, A. (2008), ‘Respiratory rate: the neglected vital sign’, *Medical Journal of Australia* **188**(11), 657–659.

- Crouter, S. E., LaMunion, S. R., Hibbing, P. R., Kaplan, A. S. & Bassett Jr, D. R. (2019), ‘Accuracy of the cosmed k5 portable calorimeter’, *PLoS One* **14**(12), e0226290.
- Cruz, R., Alves, D. L., Rumenig, E., Gonçalves, R., Degaki, E., Pasqua, L., Koch, S., Lima-Silva, A. E., S. Koehle, M. & Bertuzzi, R. (2020), ‘Estimation of minute ventilation by heart rate for field exercise studies’, *Scientific reports* **10**(1), 1423.
- Cummiskey, B., Schiffmiller, D., Talavage, T. M., Leverenz, L., Meyer, J. J., Adams, D. & Nauman, E. A. (2017), ‘Reliability and accuracy of helmet-mounted and head-mounted devices used to measure head accelerations’, *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology* **231**(2), 144–153.
- Daines, L., Buekers, J., Bolado, B. A., Kocks, J. W., Vaghegini, G., Poberezhets, V., Vogiatzis, I., Pinnock, H. & Vanfleteren, L. E. (2021), ‘Ers international congress 2020: highlights from the general pneumology assembly’, *ERJ Open Research* **7**(1).
- Daley, M. A., Bramble, D. M. & Carrier, D. R. (2013), ‘Impact loading and locomotor-respiratory coordination significantly influence breathing dynamics in running humans’, *PloS one* **8**(8), e70752.
- Daneshvar, D. H., Baugh, C. M., Nowinski, C. J., McKee, A. C., Stern, R. A. & Cantu, R. C. (2011), ‘Helmets and mouth guards: the role of personal equipment in preventing sport-related concussions’, *Clinics in sports medicine* **30**(1), 145.
- de Almeida e Bueno, L., Kwong, M. T., Milnthorpe, W. R., Cheng, R. & Bergmann, J. H. (2021), ‘Applying ubiquitous sensing to estimate perceived exertion based on cardiorespiratory features’, *Sports Engineering* **24**, 1–9.
- de Almeida e Bueno, L., Walls, V. C. & Bergmann, J. H. (2023), ‘Evaluating the

- potential of an oral-based bioguard to estimate heart rate using photoplethysmography’, *Biosensors* **13**(5), 533.
- De Fazio, R., Stabile, M., De Vittorio, M., Velázquez, R. & Visconti, P. (2021), ‘An overview of wearable piezoresistive and inertial sensors for respiration rate monitoring’, *Electronics* **10**(17), 2178.
- di Prampero, P. E., Fusi, S., Sepulcri, L., Morin, J.-B., Belli, A. & Antonutto, G. (2005), ‘Sprint running: a new energetic approach’, *Journal of experimental Biology* **208**(14), 2809–2816.
- DiFiori, J. P., Benjamin, H. J., Brenner, J., Gregory, A. Jayanthi, N. & Landry, G. L. and Luke, A. (2014), ‘Overuse injuries and burnout in youth sports: a position statement from the american medical society for sports medicine’, *Clinical Journal of Sport Medicine* **24**(1), 3–20.
- Dowling, B., McNally, M. P., Chaudhari, A. M. & Oñate, J. A. (2020), ‘A review of workload-monitoring considerations for baseball pitchers’, *Journal of athletic training* **55**(9), 911–917.
- Drew, M. K. & Finch, C. F. (2016), ‘The relationship between training load and injury, illness and soreness: a systematic and literature review’, *Sports medicine* **46**, 861–883.
- Duma, S. M., Manoogian, S. J., Bussone, W. R., Brolinson, P. G., Goforth, M. W., Donnenwerth, J. J., Greenwald, R. M., Chu, J. J. & Crisco, J. J. (2005), ‘Analysis of real-time head accelerations in collegiate football players’, *Clinical Journal of Sport Medicine* **15**(1), 3–8.
- Eime, R. M., Young, J. A., Harvey, J. T., Charity, M. J. & Payne, W. R. (2013), ‘A systematic review of the psychological and social benefits of participation in

- sport for children and adolescents: informing development of a conceptual model of health through sport', *International journal of behavioral nutrition and physical activity* **10**, 1–21.
- Elman, J. L. (1990), 'Finding structure in time', *Cognitive science* **14**(2), 179–211.
- Eston, R. (2012), 'Use of ratings of perceived exertion in sports', *International journal of sports physiology and performance* **7**(2), 175–182.
- Fahed, V. S., Doheny, E. P. & Lowery, M. M. (2023), Random forest classification of breathing phases from audio signals recorded using mobile devices, *in* 'INTER-SPEECH 2023', ISCA, pp. 2738–2742.
- Faulkner, J., Eston, R. G. et al. (2008), 'Perceived exertion research in the 21st century: developments, reflections and questions for the future'.
- Finch, C. F., Kemp, J. L. & Clapperton, A. J. (2015), 'The incidence and burden of hospital-treated sports-related injury in people aged 15+ years in victoria, australia, 2004–2010: a future epidemic of osteoarthritis?', *Osteoarthritis and cartilage* **23**(7), 1138–1143.
- Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P. & Dodge, C. (2001), 'A new approach to monitoring exercise training', *The Journal of Strength & Conditioning Research* **15**(1), 109–115.
- Gabbett, T. J., Jenkins, D. G. & Abernethy, B. (2011), 'Physical collisions and injury in professional rugby league match-play', *Journal of Science and Medicine in Sport* **14**(3), 210–215.
- Gabbett, T. J., Polley, C., Dwyer, D. B., Kearney, S. & Corvo, A. (2014), 'Influence of field position and phase of play on the physical demands of match-play

- in professional rugby league forwards', *Journal of Science and Medicine in Sport* **17**(5), 556–561.
- Gabler, L. F., Crandall, J. R. & Panzer, M. B. (2016), 'Assessment of kinematic brain injury metrics for predicting strain responses in diverse automotive impact conditions', *Annals of biomedical engineering* **44**, 3705–3718.
- Gadd, C. W. (1966), Use of a weighted-impulse criterion for estimating injury hazard, Technical report, SAE technical paper.
- Gavriely, N., Palti, Y. & Alroy, G. (1981), 'Spectral characteristics of normal breath sounds', *Journal of applied physiology* **50**(2), 307–314.
- Gessel, L. M., Fields, S. K., Collins, C. L., Dick, R. W. & Comstock, R. D. (2007), 'Concussions among united states high school and collegiate athletes', *Journal of athletic training* **42**(4), 495.
- Geurkink, Y., Vandewiele, G., Lievens, M., De Turck, F., Ongenae, F., Matthys, S. P., Boone, J. & Bourgois, J. G. (2019), 'Modeling the prediction of the session rating of perceived exertion in soccer: Unraveling the puzzle of predictive indicators', *International journal of sports physiology and performance* **14**(6), 841–846.
- Gholamiangonabadi, D., Kiselov, N. & Grolinger, K. (2020), 'Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection', *Ieee Access* **8**, 133982–133994.
- Gibson, P. G., McDonald, V. M. & Marks, G. B. (2010), 'Asthma in older adults', *The lancet* **376**(9743), 803–813.
- Gilgen-Ammann, R., Schweizer, T. & Wyss, T. (2019), 'Rr interval signal quality of a heart rate monitor and an ecg holter at rest and during exercise', *European journal of applied physiology* **119**(7), 1525–1532.

- Gould, T. E., Jesunathadas, M., Nazarenko, S. & Piland, S. G. (2019), Mouth protection in sports, *in* ‘Materials in sports equipment’, Elsevier, pp. 199–231.
- Grant, S., Aitchison, T., Henderson, E., Christie, J., Zare, S., Mc Murray, J. & Dargie, H. (1999), ‘A comparison of the reproducibility and the sensitivity to change of visual analogue scales, borg scales, and likert scales in normal subjects during submaximal exercise’, *Chest* **116**(5), 1208–1217.
- Greenwald, R. M., Gwin, J. T., Chu, J. J. & Crisco, J. J. (2008), ‘Head impact severity measures for evaluating mild traumatic brain injury risk exposure’, *Neurosurgery* **62**(4), 789–798.
- Grünbichler, J., Federolf, P. & Gatterer, H. (2020), ‘Workload efficiency as a new tool to describe external and internal competitive match load of a professional soccer team: A descriptive study on the relationship between pre-game training loads and relative match load’, *European journal of sport science* **20**(8), 1034–1041.
- Gutierrez, G. M., Conte, C. & Lightbourne, K. (2014), ‘The relationship between impact force, neck strength, and neurocognitive performance in soccer heading in adolescent females’, *Pediatric exercise science* **26**(1), 33–40.
- Haddad, M., Chaouachi, A., Wong, D. P., Castagna, C., Hambli, M., Hue, O. & Chamari, K. (2013), ‘Influence of fatigue, stress, muscle soreness and sleep on perceived exertion during submaximal effort’, *Physiology & behavior* **119**, 185–189.
- Hagel, B. & Meeuwisse, W. (2004), ‘Risk compensation: a “side effect” of sport injury prevention?’.
- Halsen, S. L. (2014), ‘Monitoring training load to understand fatigue in athletes’, *Sports medicine* **44**(Suppl 2), 139–147.

- Harbour, E., Lasshofer, M., Genitrini, M. & Schwameder, H. (2021), ‘Enhanced breathing pattern detection during running using wearable sensors’, *Sensors* **21**(16), 5606.
- Hardy, W. N., Mason, M. J., Foster, C. D., Shah, C. S., Kopacz, J. M., Yang, K. H., King, A. I., Bishop, J., Bey, M., Anderst, W. et al. (2007), ‘A study of the response of the human cadaver head to impact’, *Stapp car crash journal* **51**, 17.
- Harvill, J., Chatterjee, M., Khosla, S., Alam, M., Ahuja, N., Hasegawa-Johnson, M., Chestek, D. & Beiser, D. G. (2024), ‘Multimodal respiratory rate estimation from audio and video in emergency department patients’, *IEEE Journal of Translational Engineering in Health and Medicine* .
- Harvill, J., Wani, Y., Alam, M., Ahuja, N., Hasegawa-Johnson, M., Chestek, D. & Beiser, D. G. (2022), Estimation of respiratory rate from breathing audio, *in* ‘2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)’, IEEE, pp. 4599–4603.
- Hashim, H. A. & Sarah, A.-G. (2006), ‘Tooth width and arch dimensions in normal and malocclusion samples: an odontometric study’, *The journal of contemporary dental practice* **6**(2), 36–51.
- Haste, P., e Bueno, L. d. A., Jérusalem, A. & Bergmann, J. (2025), ‘Performance of current tools used for on-the-day assessment and diagnosis of mild traumatic brain injury in sport: a systematic review’, *BMJ Open Sport & Exercise Medicine* **11**(1).
- Hettiarachchi, C., Kodithuwakku, J., Manamperi, B., Ifham, A. & Silva, P. (2019), A wearable system to analyze the human arm for predicting injuries due to throwing, *in* ‘2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)’, IEEE, pp. 3297–3301.

- Higham, D. G., Pyne, D. B., Anson, J. M. & Eddy, A. (2013), ‘Physiological, anthropometric, and performance characteristics of rugby sevens players’, *International journal of sports physiology and performance* **8**(1), 19–27.
- Hills, A. P., Andersen, L. B. & Byrne, N. M. (2011), ‘Physical activity and obesity in children’, *British journal of sports medicine* **45**(11), 866–870.
- Hochreiter, S. & Schmidhuber, J. (1997), ‘Long short-term memory’, *Neural Computation MIT-Press* .
- Hou, X. (2023), Robust localization with wearable sensors, PhD thesis, University of Oxford.
- Houssein, A., Ge, D., Gastinger, S., Dumond, R. & Prioux, J. (2021), ‘A novel algorithm for minute ventilation estimation in remote health monitoring with magnetometer plethysmography’, *Computers in Biology and Medicine* **130**, 104189.
- Hu, C., Kandappu, T., Liu, Y., Mascolo, C. & Ma, D. (2024), ‘Breathpro: Monitoring breathing mode during running with earables’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **8**(2), 1–25.
- Hulin, B. T., Gabbett, T. J., Johnston, R. D. & Jenkins, D. G. (2017), ‘Wearable microtechnology can accurately identify collision events during professional rugby league match-play’, *Journal of Science and Medicine in Sport* **20**(7), 638–642.
- Hussain, T., Ullah, S., Fernández-García, R. & Gil, I. (2023), ‘Wearable sensors for respiration monitoring: A review’, *Sensors* **23**(17), 7518.
- Hutchinson, J., Kaiser, M. J. & Lankarani, H. M. (1998), ‘The head injury criterion (hic) functional’, *Applied mathematics and computation* **96**(1), 1–16.
- Iorio, O. C., Coutu, F.-A., Malaeb, D. & Ross, B. A. (2024), ‘Feasibility, functionality,

- and user experience with wearable technologies for acute exacerbation monitoring in patients with severe copd’, *Frontiers in Signal Processing* **4**, 1362754.
- Jadischke, R., Viano, D. C., Dau, N., King, A. I. & McCarthy, J. (2013), ‘On the accuracy of the head impact telemetry (hit) system used in football helmets’, *Journal of biomechanics* **46**(13), 2310–2315.
- Ji, S. & Zhao, W. (2015), ‘A pre-computed brain response atlas for instantaneous strain estimation in contact sports’, *Annals of biomedical engineering* **43**(8), 1877–1895.
- Ji, S., Zhao, W., Li, Z. & McAllister, T. W. (2014), ‘Head impact accelerations for brain strain-related responses in contact sports: a model-based investigation’, *Biomechanics and modeling in mechanobiology* **13**(5), 1121–1136.
- Jia, Z., Huth, H., Teoh, W. Q., Xu, S., Wood, B. & Tse, Z. T. H. (2025), ‘State of the art review of wearable devices for respiratory monitoring’, *IEEE Access* **13**, 18178–18190.
- Jones, C. M., Griffiths, P. C. & Mellalieu, S. D. (2017), ‘Training load and fatigue marker associations with injury and illness: a systematic review of longitudinal studies’, *Sports medicine* **47**, 943–974.
- Karacocuk, G., Höflinger, F., Zhang, R., Reindl, L. M., Laufer, B., Möller, K., Röell, M. & Zdzieblik, D. (2019), ‘Inertial sensor-based respiration analysis’, *IEEE Transactions on Instrumentation and Measurement* **68**(11), 4268–4275.
- Keaney, L. C., Kilding, A. E., Merien, F. & Dulson, D. K. (2018), ‘The impact of sport related stressors on immunity and illness risk in team-sport athletes’, *Journal of science and medicine in sport* **21**(12), 1192–1199.

- Kim, D., Lee, J., Park, M. K. & Ko, S. H. (2024), ‘Recent developments in wearable breath sensors for healthcare monitoring’, *Communications Materials* **5**(1), 41.
- Kim, J.-H., Roberge, R., Powell, J., Shafer, A. & Williams, W. J. (2013), ‘Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the zephyr bioharnesstm’, *International journal of sports medicine* **34**(06), 497–501.
- Kim, K. B. & Baek, H. J. (2023), ‘Photoplethysmography in wearable devices: a comprehensive review of technological advances, current challenges, and future directions’, *Electronics* **12**(13), 2923.
- Kimpara, H. & Iwamoto, M. (2012), ‘Mild traumatic brain injury predictors based on angular accelerations during impacts’, *Annals of biomedical engineering* **40**, 114–126.
- Kleiven, S. (2003), ‘Influence of impact direction on the human head in prediction of subdural hematoma’, *Journal of neurotrauma* **20**(4), 365–379.
- Kleiven, S. (2007), Predictors for traumatic brain injuries evaluated through accident reconstructions, Technical report, SAE Technical Paper.
- Knapik, J. J., Marshall, S. W., Lee, R. B., Darakjy, S. S., Jones, S. B., Mitchener, T. A., dela Cruz, G. G. & Jones, B. H. (2007), ‘Mouthguards in sport activities history, physical properties and injury prevention effectiveness’, *Sports medicine* **37**, 117–144.
- Krustrup, P., Mohr, M., Amstrup, T., Rysgaard, T., Johansen, J., Steensberg, A., Pedersen, P. K. & Bangsbo, J. (2003), ‘The yo-yo intermittent recovery test: physiological response, reliability, and validity’, *Medicine & Science in Sports & Exercise* **35**(4), 697–705.

- Krustrup, P., Mohr, M., Nybo, L., Jensen, J. M., Nielsen, J. J. & Bangsbo, J. (2006), ‘The yo-yo ir2 test: physiological response, reliability, and application to elite soccer’, *Medicine & Science in Sports & Exercise* **38**(9), 1666–1673.
- Kujala, U. M., Taimela, S., Antti-Poika, I., Orava, S., Tuominen, R. & Myllynen, P. (1995), ‘Acute injuries in soccer, ice hockey, volleyball, basketball, judo, and karate: analysis of national registry data’, *Bmj* **311**(7018), 1465–1468.
- Kumar, A., Mitra, V., Oliver, C., Ullal, A., Biddulph, M. & Mance, I. (2021), Estimating respiratory rate from breath audio obtained through wearable microphones, in ‘2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)’, IEEE, pp. 7310–7315.
- Kuo, C., Wu, L., Loza, J., Senif, D., Anderson, S. C. & Camarillo, D. B. (2018), ‘Comparison of video-based and sensor-based head impact exposure’, *PloS one* **13**(6), e0199238.
- Kupperman, N. & Hertel, J. (2020), ‘Global positioning system–derived workload metrics and injury risk in team-based field sports: a systematic review’, *Journal of athletic training* **55**(9), 931–943.
- Kusche, R., Klimach, P., Malhotra, A., Kaufmann, S. & Ryschka, M. (2015), ‘An in-ear pulse wave velocity measurement system using heart sounds as time reference’, *Current Directions in Biomedical Engineering* **1**(1), 366–370.
- Lanata, A., Scilingo, E. P., Nardini, E., Loriga, G., Paradiso, R. & De-Rossi, D. (2009), ‘Comparative evaluation of susceptibility to motion artifact in different wearable systems for monitoring respiratory rate’, *IEEE Transactions on Information Technology in Biomedicine* **14**(2), 378–386.
- Langlois, J. A., Rutland-Brown, W. & Wald, M. M. (2006), ‘The epidemiology and

- impact of traumatic brain injury: a brief overview', *The Journal of head trauma rehabilitation* **21**(5), 375–378.
- LaRocca, D., James, K. A., Rosenberg, S., Ma, M. & Brooks-Russell, A. (2023), 'Team sports participation, depression, and suicidal ideation in lesbian, gay, bisexual, transgender, and questioning adolescents', *Psychology in the Schools* **60**(4), 902–911.
- LaRoche, A. A., Nelson, L. D., Connelly, P. K., Walter, K. D. & McCrea, M. A. (2016), 'Sport-related concussion reporting and state legislative effects', *Clinical journal of sport medicine* **26**(1), 33–39.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015), 'Deep learning', *nature* **521**(7553), 436–444.
- Li, R. T., Kling, S. R., Salata, M. J., Cupp, S. A., Sheehan, J. & Voos, J. E. (2016), 'Wearable performance devices in sports medicine', *Sports health* **8**(1), 74–78.
- Li, Z., Liu, F., Yang, W., Peng, S. & Zhou, J. (2021), 'A survey of convolutional neural networks: analysis, applications, and prospects', *IEEE transactions on neural networks and learning systems* **33**(12), 6999–7019.
- Lim, C., Kim, J., Kim, J., Kang, B.-G. & Nam, Y. (2022), 'Estimation of respiratory rate in various environments using microphones embedded in face masks', *The Journal of Supercomputing* **78**(17), 19228–19245.
- Ling, H., Hardy, J. & Zetterberg, H. (2015), 'Neurological consequences of traumatic brain injuries in sports', *Molecular and Cellular Neuroscience* **66**, 114–122.
- Liu, H., Allen, J., Zheng, D. & Chen, F. (2019), 'Recent development of respiratory rate measurement technologies', *Physiological measurement* **40**(7), 07TR01.

- Liu, S., Gao, R. X. & Freedson, P. S. (2011), Non-invasive respiration and ventilation prediction using a single abdominal sensor belt, *in* ‘2011 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)’, IEEE, pp. 1–5.
- Liu, Y., Butkow, K.-J., Stuchbury-Wass, J., Pullin, A., Ma, D. & Mascolo, C. (2025), Respear: Earable-based robust respiratory rate monitoring, *in* ‘2025 IEEE International Conference on Pervasive Computing and Communications (PerCom)’, IEEE, pp. 67–77.
- Liu, Y., Shukla, D., Newman, H. & Zhu, Y. (2022), ‘Soft wearable sensors for monitoring symptoms of covid-19 and other respiratory diseases: a review’, *Progress in Biomedical Engineering* **4**(1), 012001.
- Löllgen, H. & Leyk, D. (2018), ‘Exercise testing in sports medicine’, *Deutsches Ärzteblatt International* **115**(24), 409.
- Luke, A., Lazaro, R. M., Bergeron, M. F., Keyser, L., Benjamin, H., Brenner, J., d’Hemecourt, P., Grady, M., Philpott, J. & Smith, A. (2011), ‘Sports-related injuries in youth athletes: is overscheduling a risk factor?’, *Clinical journal of sport medicine* **21**(4), 307–314.
- Lutz, J., Memmert, D., Raabe, D., Dornberger, R. & Donath, L. (2020), ‘Wearables for integrative performance and tactic analyses: Opportunities, challenges, and future directions’, *International journal of environmental research and public health* **17**(1), 59.
- MacLeod, S. J., Hagan, C., Egaña, M., Davis, J. & Drake, D. (2018), ‘The use of microtechnology to monitor collision performance in professional rugby union’, *International journal of sports physiology and performance* **13**(8), 1075–1082.
- Marcora, S. (2010), ‘Counterpoint: afferent feedback from fatigued locomotor muscles

- is not an important determinant of endurance exercise performance’, *Journal of applied physiology* **108**(2), 454–456.
- Maron, B. J., Gohman, T. E. & Aepli, D. (1998), ‘Prevalence of sudden cardiac death during competitive sports activities in minnesota high school athletes’, *Journal of the American College of Cardiology* **32**(7), 1881–1884.
- Marshall, D. A., Lopatina, E., Lacny, S. & Emery, C. A. (2016), ‘Economic impact study: neuromuscular training reduces the burden of injuries and costs compared to standard warm-up in youth soccer’, *British journal of sports medicine* **50**(22), 1388–1393.
- Martin, A. & Voix, J. (2017), ‘In-ear audio wearable: Measurement of heart and breathing rates for health and safety monitoring’, *IEEE Transactions on Biomedical Engineering* **65**(6), 1256–1263.
- Massaroni, C., Nicolò, A., Lo Presti, D., Sacchetti, M., Silvestri, S. & Schena, E. (2019), ‘Contact-based methods for measuring respiratory rate’, *Sensors* **19**(4), 908.
- Massone, P., Angelucci, A. & Aliverti, A. (2024), On the measurement of tidal volume with wearable devices, *in* ‘International Workshop on Engineering Methodologies for Medicine and Sport’, Springer, pp. 117–128.
- McAllister, T. W., Ford, J. C., Ji, S., Beckwith, J. G., Flashman, L. A., Paulsen, K. & Greenwald, R. M. (2012), ‘Maximum principal strain and strain rate associated with concussion diagnosis correlates with changes in corpus callosum white matter indices’, *Annals of biomedical engineering* **40**, 127–140.
- McCrea, M., Hammeke, T., Olsen, G., Leo, P. & Guskiewicz, K. (2004), ‘Unreported concussion in high school football players: implications for prevention’, *Clinical journal of sport medicine* **14**(1), 13–17.

- Merrell, A. J., Christensen, W. F., Seeley, M. K., Bowden, A. E. & Fullwood, D. T. (2017), ‘Nano-composite foam sensor system in football helmets’, *Annals of biomedical engineering* **45**, 2742–2749.
- Mihalik, J. P., Chandran, A., Powell, J. R., Roby, P. R., Guskiewicz, K. M., Stemper, B. D., Shah, A. S., Rowson, S., Duma, S., Harezlak, J. et al. (2020), ‘Do head injury biomechanics predict concussion clinical recovery in college american football players?’, *Annals of biomedical engineering* **48**, 2555–2565.
- Miller, M. R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., Crapo, R., Enright, P., Van Der Grinten, C., Gustafsson, P. et al. (2005), ‘Standardisation of spirometry’, *European respiratory journal* **26**(2), 319–338.
- Mitchell, T. M. & Mitchell, T. M. (1997), *Machine learning*, Vol. 1, McGraw-hill New York.
- Mohktar, M. S., Redmond, S. J., Antoniadis, N. C., Rochford, P. D., Pretto, J. J., Basilakis, J., Lovell, N. H. & McDonald, C. F. (2015), ‘Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data’, *Artificial intelligence in medicine* **63**(1), 51–59.
- Monaco, V. & Stefanini, C. (2021), ‘Assessing the tidal volume through wearables: a scoping review’, *Sensors* **21**(12), 4124.
- Muniz-Pardos, B., Angeloudis, K., Guppy, F. M., Keramitsoglou, I., Sutehall, S., Bosch, A., Tanisawa, K., Hosokawa, Y., Ash, G. I., Schobersberger, W. et al. (2021), ‘Wearable and telemedicine innovations for olympic events and elite sport’, *Journal of Sports Medicine and Physical Fitness* **61**(8), 1061–1072.
- Murray, N. B., Gabbett, T. J., Townshend, A. D. & Blanch, P. (2017), ‘Calculating acute: chronic workload ratios using exponentially weighted moving averages pro-

- vides a more sensitive indicator of injury likelihood than rolling averages’, *British Journal of Sports Medicine* **51**(9), 749–754.
- Nagakumar, P., Do, W., Carr, O., Saglani, S. & Fleming, L. (2024), Contactless bedside monitoring using the albus home research device (rd) to predict severe asthma attacks in children, *in* ‘C26. NOVEL TECHNOLOGIES AND IMAGING APPROACHES IN PEDIATRIC RESPIRATORY MEDICINE’, American Thoracic Society, pp. A5167–A5167.
- Nam, Y., Reyes, B. A. & Chon, K. H. (2015), ‘Estimation of respiratory rates using the built-in microphone of a smartphone or headset’, *IEEE journal of biomedical and health informatics* **20**(6), 1493–1501.
- Namjoshi, D. R., Good, C., Cheng, W. H., Panenka, W., Richards, D., Cripton, P. A. & Wellington, C. L. (2013), ‘Towards clinical management of traumatic brain injury: a review of models and mechanisms from a biomechanical perspective’, *Disease models & mechanisms* **6**(6), 1325–1338.
- Naranjo-Hernández, D., Talaminos-Barroso, A., Reina-Tosina, J., Roa, L. M., Barbarov-Rostan, G., Cejudo-Ramos, P., Márquez-Martín, E. & Ortega-Ruiz, F. (2018), ‘Smart vest for respiratory rate monitoring of copd patients based on non-contact capacitive sensing’, *Sensors* **18**(7), 2144.
- Naughton, M., Jones, B., Hendricks, S., King, D., Murphy, A. & Cummins, C. (2020), ‘Quantifying the collision dose in rugby league: a systematic review, meta-analysis, and critical analysis’, *Sports medicine-open* **6**, 1–28.
- Naughton, M., Miller, J. & Slater, G. J. (2018), ‘Impact-induced muscle damage and contact sports: etiology, effects on neuromuscular function and recovery, and the modulating effects of adaptation and recovery strategies’, *International Journal of Sports Physiology and Performance* **13**(8), 962–969.

- Nauman, E. A., Breedlove, K. M., Breedlove, E. L., Talavage, T. M., Robinson, M. E. & Leverenz, L. J. (2015), ‘Post-season neurophysiological deficits assessed by impact and fmri in athletes competing in american football’, *Developmental neuropsychology* **40**(2), 85–91.
- Newman, J. A. (1986), A generalized acceleration model for brain injury threshold (gambit), *in* ‘Proceedings of International Conference on the Biomechanics of Impact, 1986’, pp. 121–131.
- Newman, J. A. & Shewchenko, N. (2000), A proposed new biomechanical head injury assessment function-the maximum power index, Technical report, SAE Technical Paper.
- Nicolò, A., Marcora, S. M. & Sacchetti, M. (2016), ‘Respiratory frequency is strongly associated with perceived exertion during time trials of different duration’, *Journal of sports sciences* **34**(13), 1199–1206.
- Nicolò, A., Massaroni, C. & Passfield, L. (2017), ‘Respiratory frequency during exercise: the neglected physiological measure’, *Frontiers in physiology* **8**, 922.
- Nicolò, A., Massaroni, C., Schena, E. & Sacchetti, M. (2020), ‘The importance of respiratory rate monitoring: From healthcare to sport and exercise’, *Sensors* **20**(21), 6396.
- Nicolò, A., Montini, M., Girardi, M., Felici, F., Bazzucchi, I. & Sacchetti, M. (2020), ‘Respiratory frequency as a marker of physical effort during high-intensity interval training in soccer players’, *International journal of sports physiology and performance* **15**(1), 73–80.
- O’Connor, K. L., Rowson, S., Duma, S. M. & Broglio, S. P. (2017), ‘Head-impact–measurement devices: a systematic review’, *Journal of athletic training* **52**(3), 206–227.

- Oliva-Lozano, J. M., Muyor, J. M., Fortes, V. & McLaren, S. J. (2021), ‘Decomposing the variability of match physical performance in professional soccer: Implications for monitoring individuals’, *European Journal of Sport Science* **21**(11), 1588–1596.
- O’Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L. et al. (2019), ‘Kerastuner’, <https://github.com/keras-team/keras-tuner>.
- Ommaya, A. K., Yarnell, P., Hirsch, A. E. & Harris, E. H. (1967), Scaling of experimental data on cerebral concussion in sub-human primates to concussion threshold for man, Technical report, SAE Technical Paper.
- Pedregosa, F. (2011), ‘Scikit-learn: Machine learning in python fabian’, *Journal of machine learning research* **12**, 2825.
- Pellman, EJ, V. D. T. A. C. I. & Waeckerle, J. (2003), ‘Concussion in professional football: reconstruction of game impacts and injuries’, *neurosurgery* **53**, 799–812.
- Perez-Suarez, I., Martin-Rincon, M., Gonzalez-Henriquez, J. J., Fezzardi, C., Perez-Regalado, S., Galvan-Alvarez, V., Juan-Habib, J. W., Morales-Alamo, D. & Calbet, J. A. (2018), ‘Accuracy and precision of the cosmed k5 portable analyser’, *Frontiers in physiology* **9**, 1764.
- Perri, E., Simonelli, C., Rossi, A., Trecroci, A., Alberti, G. & Iaia, F. M. (2021), ‘Relationship between wellness index and internal training load in soccer: application of a machine learning model’, *International Journal of Sports Physiology and Performance* **16**(5), 695–703.
- Peterson, A. J., Foster, C., Yee, N. B., Stamatopolous, C. C., Giotis, P., Wright, G. A., Doberstein, S. T. & Porcari, J. P. (2014), ‘Acoustic analysis of breath sounds as a surrogate for gas exchange thresholds’, *Journal of Sports Science & Medicine* **13**(1), 219.

- Pharr, J. R., Lough, N. L. & Terencio, M. A. (2019), ‘Health and sociodemographic differences between individual and team sport participants’, *Sports* **7**(6), 150.
- Pirscoveanu, C.-I. & Oliveira, A. S. (2024), ‘Prediction of instantaneous perceived effort during outdoor running using accelerometry and machine learning’, *European Journal of Applied Physiology* **124**(3), 963–973.
- Popescu, M.-C., Balas, V. E., Perescu-Popescu, L. & Mastorakis, N. (2009), ‘Multilayer perceptron and neural networks’, *WSEAS Transactions on Circuits and Systems* **8**(7), 579–588.
- Pressler, G. A., Mansfield, J. P., Pasterkamp, H. & Wodicka, G. R. (2004), ‘Detection of respiratory sounds at the external ear’, *IEEE transactions on biomedical engineering* **51**(12), 2089–2096.
- Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S.-Y. & Sainath, T. (2019), ‘Deep learning for audio signal processing’, *IEEE Journal of Selected Topics in Signal Processing* **13**(2), 206–219.
- Ragaišienė, G., Kibarskytė, R., Gauronskaitė, R., Giedraitytė, M., Dapšauskaitė, A., Kasiulevičius, V. & Danila, E. (2019), ‘Diagnosing copd in primary care: what has real life practice got to do with guidelines?’, *Multidisciplinary respiratory medicine* **14**, 1–7.
- Raglin, J., Sawamura, S., Alexiou, S., Hassmen, P. & Kenttä, G. (2000), ‘Training practices and staleness in 13–18-year-old swimmers: A cross-cultural study’, *Pediatric Exercise Science* **12**(1), 61–70.
- Rago, V., Brito, J., Figueiredo, P., Costa, J., Barreira, D., Krstrup, P. & Rebelo, A. (2020), ‘Methods to collect and interpret external training load using microtechnology incorporating gps in professional football: a systematic review’, *Research in Sports Medicine* **28**(3), 437–458.

- Rahman, M. S., Chowdhury, S., Rasheduzzaman, M. & Doulah, A. B. M. S. U. (2024), ‘Artificial intelligence-based algorithms and healthcare applications of respiratory inductance plethysmography: A systematic review’, *Algorithms* **17**(6), 261.
- Raicevic, N., Ivanovic, M., Belicev, P. & Petrovic, J. (2016), Monitoring of respiratory volumes by an long period grating sensor of bending, *in* ‘Journal of Physics: Conference Series’, Vol. 682, IOP Publishing, p. 012008.
- Rampinini, E., Sassi, A., Azzalin, A., Castagna, C., Menaspa, P., Carlomagno, D. & Impellizzeri, F. M. (2010), ‘Physiological determinants of yo-yo intermittent recovery tests in male soccer players’, *European journal of applied physiology* **108**(2), 401–409.
- Ren, Y., Zheng, Z., Liu, H., Chen, Y., Li, H. & Wang, C. (2021), Breathing sound-based exercise intensity monitoring via smartphones, *in* ‘2021 international conference on computer communications and networks (ICCCN)’, IEEE, pp. 1–10.
- Retory, Y., Niedzialkowski, P., De Picciotto, C., Bonay, M. & Petitjean, M. (2016), ‘New respiratory inductive plethysmography (rip) method for evaluating ventilatory adaptation during mild physical activities’, *PLoS One* **11**(3), e0151983.
- Rich, A. M., Filben, T. M., Miller, L. E., Tomblin, B. T., Van Gorkom, A. R., Hurst, M. A., Barnard, R. T., Kohn, D. S., Urban, J. E. & Stitzel, J. D. (2019), ‘Development, validation and pilot field deployment of a custom mouthpiece for head impact measurement’, *Annals of biomedical engineering* **47**, 2109–2121.
- Rico-González, M., Los Arcos, A., Clemente, F. M., Rojas-Valverde, D. & Pino-Ortega, J. (2020), ‘Accuracy and reliability of local positioning systems for measuring sport movement patterns in stadium-scale: A systematic review’, *Applied sciences* **10**(17), 5994.

- Romano, C., Nicolò, A., Innocenti, L., Bravi, M., Miccinilli, S., Sterzi, S., Sacchetti, M., Schena, E. & Massaroni, C. (2023), ‘Respiratory rate estimation during walking and running using breathing sounds recorded with a microphone’, *Biosensors* **13**(6), 637.
- Rossi, A., Perri, E., Pappalardo, L., Cintia, P. & Iaia, F. M. (2019), ‘Relationship between external and internal workloads in elite soccer players: comparison between rate of perceived exertion and training load’, *Applied sciences* **9**(23), 5174.
- Rowson, S. & Duma, S. M. (2013), ‘Brain injury prediction: assessing the combined probability of concussion using linear and rotational head acceleration’, *Annals of biomedical engineering* **41**, 873–882.
- Ryan, J. L., Pracht, E. E. & Orban, B. L. (2019), ‘Inpatient and emergency department costs from sports injuries among youth aged 5–18 years’, *BMJ open sport & exercise medicine* **5**(1), e000491.
- Sarkar, M., Madabhavi, I., Niranjana, N. & Dogra, M. (2015), ‘Auscultation of the respiratory system’, *Annals of thoracic medicine* **10**(3), 158–168.
- Schaffarczyk, M., Rogers, B., Reer, R. & Gronwald, T. (2022), ‘Validity of the polar h10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women’, *Sensors* **22**(17), 6536.
- Schmitz, N., Kruse, J. & Kugler, J. (2004), ‘The association between physical exercises and health-related quality of life in subjects with mental disorders: results from a cross-sectional survey’, *Preventive medicine* **39**(6), 1200–1207.
- Schneider, C., Hanakam, F., Wiewelhove, T., Döweling, A., Kellmann, M., Meyer, T., Pfeiffer, M. & Ferrauti, A. (2018), ‘Heart rate monitoring in team sports—a conceptual framework for contextualizing heart rate measures for training and recovery prescription’, *Frontiers in physiology* **9**, 639.

- Schuster, M. & Paliwal, K. K. (1997), ‘Bidirectional recurrent neural networks’, *IEEE transactions on Signal Processing* **45**(11), 2673–2681.
- Schweiger, D., Critcher, S., Freeborn, T. J. & Hibberd, E. (2020), Throwing event detection using acceleration magnitude collected with wrist-worn sensors, *in* ‘2020 SoutheastCon’, IEEE, pp. 1–7.
- Seshadri, D. R., Li, R. T., Voos, J. E., Rowbottom, J. R., Alfes, C. M., Zorman, C. A. & Drummond, C. K. (2019), ‘Wearable sensors for monitoring the physiological and biochemical profile of the athlete’, *NPJ digital medicine* **2**(1), 72.
- Seshadri, D. R., Magliato, S., Voos, J. E. & Drummond, C. (2019), ‘Clinical translation of biomedical sensors for sports medicine’, *Journal of medical engineering & technology* **43**(1), 66–81.
- Seshadri, D. R., Rowbottom, J. R., Drummond, C., Voos, J. E. & Craker, J. (2016), A review of wearable technology: Moving beyond the hype: From need through sensor implementation, *in* ‘2016 8th Cairo International Biomedical Engineering Conference (CIBEC)’, IEEE, pp. 52–55.
- Shi, L., Han, Y., Huang, H., Davidsson, J. & Thomson, R. (2020), ‘Evaluation of injury thresholds for predicting severe head injuries in vulnerable road users resulting from ground impact via detailed accident reconstructions’, *Biomechanics and modeling in mechanobiology* **19**, 1845–1863.
- Siegmund, G. P., Guskiewicz, K. M., Marshall, S. W., DeMarco, A. L. & Bonin, S. J. (2016), ‘Laboratory validation of two wearable sensor systems for measuring head impact severity in football players’, *Annals of biomedical engineering* **44**, 1257–1274.
- Slobounov, S. (2008), ‘Fatigue-related injuries in athletes’, *Injuries in athletics: causes and consequences* pp. 77–95.

- Soliman, M. M., Ganti, V. G. & Inan, O. T. (2022), ‘Toward wearable estimation of tidal volume via electrocardiogram and seismocardiogram signals’, *IEEE sensors journal* **22**(18), 18093–18103.
- Souhail, H., Castagna, C., Yahmed Mohamed, H., Younes, H., Chamari, K. et al. (2010), ‘Direct validity of the yo-yo intermittent recovery test in young team handball players’, *The Journal of Strength & Conditioning Research* **24**(2), 465–470.
- Sousa, A. M., Pinho, A. C. & Piedade, A. P. (2021), ‘Mechanical properties of 3d printed mouthguards: Influence of layer height and device thickness’, *Materials & Design* **203**, 109624.
- Stagno, K. M., Thatcher, R. & Van Someren, K. A. (2007), ‘A modified trimp to quantify the in-season training load of team sport players’, *Journal of sports sciences* **25**(6), 629–634.
- Stone, T., Stone, N., Roy, N., Melton, W., Jackson, J. B. & Nelakuditi, S. (2018), ‘On smart soccer ball as a head impact sensor’, *IEEE Transactions on Instrumentation and Measurement* **68**(8), 2979–2987.
- Stretch, R. A. (1995), ‘The seasonal incidence and nature of injuries in schoolboy cricketers’, *South African Medical Journal* **85**(11).
- Tahira, S. (2022), ‘The association between sports participation and mental health across the lifespan’, *International Journal of Sport Studies for Health* **5**(2), 1–8.
- Takhounts, E. G., Craig, M. J., Moorhouse, K., McFadden, J. & Hasija, V. (2013), Development of brain injury criteria (bric), Technical report, SAE Technical Paper.
- Takhounts, E. G., Hasija, V., Ridella, S. A., Rowson, S. & Duma, S. M. (2011), Kinematic rotational brain injury criterion (bric), in ‘Proceedings of the 22nd enhanced safety of vehicles conference. Paper’, number 11-0263, pp. 1–10.

- Tanabe, G., Churei, H., Wada, T., Takahashi, H., Uo, M. & Ueno, T. (2020), ‘The influence of temperature on sheet lamination process when fabricating mouthguard on dental thermoforming machine’, *Journal of Oral Science* **62**(1), 23–27.
- Teličko, J. & Jakovičs, A. (2024), Esp32 based breathing rate monitoring system for sport activities, in ‘2024 IEEE 65th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTU CON)’, IEEE, pp. 1–6.
- Theodoropoulos, J. S., Bettel, J. & Kosy, J. D. (2020), ‘The use of gps and inertial devices for player monitoring in team sports: A review of current and future applications’, *Orthopedic reviews* **12**(1).
- Thornton, H. R., Delaney, J. A., Duthie, G. M. & Dascombe, B. J. (2019), ‘Developing athlete monitoring systems in team sports: data analysis and visualization’, *International journal of sports physiology and performance* **14**(6), 698–705.
- Thorpe, R. T., Atkinson, G., Drust, B. & Gregson, W. (2017), ‘Monitoring fatigue status in elite team-sport athletes: implications for practice’, *International journal of sports physiology and performance* **12**(s2), S2–27.
- Umberto, M. (2018), ‘Applied deep learning: A case-based approach to understanding deep neural networks’, *Appl. Deep Learn. Berkeley, California, USA Apress* pp. 185–216.
- Vallance, E., Sutton-Charani, N., Imoussaten, A., Montmain, J. & Perrey, S. (2020), ‘Combining internal-and external-training-loads to predict non-contact injuries in soccer’, *Applied Sciences* **10**(15), 5261.
- Vandewiele, G., Geurkink, Y., Lievens, M., Ongenae, F., De Turck, F. & Boone, J. (2017), Enabling training personalization by predicting the session rate of per-

- ceived exertion (srpe), in ‘Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2017 Workshop’, pp. 1–12.
- Vanegas, E., Igual, R. & Plaza, I. (2020), ‘Sensing systems for respiration monitoring: A technical systematic review’, *Sensors* **20**(18), 5446.
- VanItallie, T. B. (2019), ‘Traumatic brain injury (tbi) in collision sports: Possible mechanisms of transformation into chronic traumatic encephalopathy (cte)’, *Metabolism* **100**, 153943.
- Vasquez-Bonilla, A. A., Camacho-Cardenosa, A., Timón, R., Martínez-Guardado, I., Camacho-Cardenosa, M. & Olcina, G. (2021), ‘Muscle oxygen desaturation and re-saturation capacity limits in repeated sprint ability performance in women soccer players: a new physiological interpretation’, *International journal of environmental research and public health* **18**(7), 3484.
- Versace, J. (1971), ‘A review of the severity index’, *15th Stapp Car Crash Conference* .
- Vicente, B. A., Sebastiao, R. & Sencadas, V. (2024), ‘Wearable devices for respiratory monitoring’, *Advanced Functional Materials* **34**(45), 2404348.
- Vitazkova, D., Foltan, E., Kosnacova, H., Micjan, M., Donoval, M., Kuzma, A., Kopani, M. & Vavrinsky, E. (2024), ‘Advances in respiratory monitoring: A comprehensive review of wearable and remote technologies’, *Biosensors* **14**(2), 90.
- Vriend, I., Gouttebarga, V., Finch, C. F., Van Mechelen, W. & Verhagen, E. A. (2017), ‘Intervention strategies used in sport injury prevention studies: a systematic review identifying studies applying the haddon matrix’, *Sports medicine* **47**, 2027–2043.
- Warburton, D. E., Nicol, C. W. & Bredin, S. S. (2006), ‘Health benefits of physical activity: the evidence’, *Cmaj* **174**(6), 801–809.

- Wellman, A. D., Coad, S. C., Goulet, G. C. & McLellan, C. P. (2017), ‘Quantification of accelerometer derived impacts associated with competitive games in national collegiate athletic association division i college football players’, *The Journal of Strength & Conditioning Research* **31**(2), 330–338.
- West, S. W., Williams, S., Cazzola, D., Kemp, S., Cross, M. J. & Stokes, K. A. (2021), ‘Training load and injury risk in elite rugby union: the largest investigation to date’, *International journal of sports medicine* **42**(08), 731–739.
- Wientjes, C. J. (1992), ‘Respiration in psychophysiology: methods and applications’, *Biological psychology* **34**(2-3), 179–203.
- Williamson, I. & Goodman, D. (2006), ‘Converging evidence for the under-reporting of concussions in youth ice hockey’, *British journal of sports medicine* **40**(2), 128–132.
- Winkert, K., Kirsten, J., Dreyhaupt, J., Steinacker, J. M. & Treff, G. (2020), ‘The cosmed k5 in breath-by-breath and mixing chamber mode at low to high intensities.’, *Medicine and science in sports and exercise* **52**(5), 1153–1162.
- Worsey, M. T., Jones, B. S., Cervantes, A., Chauvet, S. P., Thiel, D. V. & Espinosa, H. G. (2020), ‘Assessment of head impacts and muscle activity in soccer using a t3 inertial sensor and a portable electromyography (emg) system: a preliminary study’, *Electronics* **9**(5), 834.
- Wu, L. C., Kuo, C., Loza, J., Kurt, M., Laksari, K., Yanez, L. Z., Senif, D., Anderson, S. C., Miller, L. E., Urban, J. E. et al. (2017), ‘Detection of american football head impacts using biomechanical features and support vector machine classification’, *Scientific reports* **8**(1), 855.
- Xue, B., Shi, W., Chotirmall, S. H., Koh, V. C. A., Ang, Y. Y., Tan, R. X. & Ser, W.

- (2022), ‘Distance-based detection of cough, wheeze, and breath sounds on wearable devices’, *Sensors* **22**(6), 2167.
- Yadollahi, A., Montazeri, A., Azarbarzin, A. & Moussavi, Z. (2013), ‘Respiratory flow–sound relationship during both wakefulness and sleep and its variation in relation to sleep apnea’, *Annals of biomedical engineering* **41**, 537–546.
- Yañez, A. M., Guerrero, D., de Alejo, R. P., Garcia-Rio, F., Alvarez-Sala, J. L., Calle-Rubio, M., de Molina, R. M., Falcones, M. V., Ussetti, P., Sauleda, J. et al. (2012), ‘Monitoring breathing rate at home allows early identification of copd exacerbations’, *Chest* **142**(6), 1524–1529.
- Yang, K. H., Hu, J., White, N. A., King, A. I., Chou, C. C. & Prasad, P. (2006), ‘Development of numerical models for injury biomechanics research: a review of 50 years of publications in the stapp car crash conference’, *50th Stapp Car Crash Conference* .
- Yedida, R. & Saha, S. (2021), ‘Beginning with machine learning: a comprehensive primer’, *The European Physical Journal Special Topics* **230**, 2363–2444.
- Yin, Z., Yang, Y., Hu, C., Li, J., Qin, B. & Yang, X. (2024), ‘Wearable respiratory sensors for health monitoring’, *NPG Asia Materials* **16**(1), 8.
- Zhang, Z., Jin, W., Huang, D. & Sun, Z. (2025), ‘A novel audio-perception-based algorithm for physiological monitoring’, *Sensors* **25**(12), 3582.
- Zhang, Z., Zheng, J., Wu, H., Wang, W., Wang, B. & Liu, H. (2012), ‘Development of a respiratory inductive plethysmography module supporting multiple sensors for wearable systems’, *Sensors* **12**(10), 13167–13184.
- Zhao, H., Xu, Y., Wu, Y., Ma, Z., Ding, Z. & Sun, Y. (2023), ‘Modeling of the rating

of perceived exertion based on heart rate using machine learning methods’, *Anais da Academia Brasileira de Ciências* **95**, e20201723.

Zhou, B., Baucells Costa, A. & Lukowicz, P. (2020), ‘Accurate spirometry with integrated barometric sensors in face-worn garments’, *Sensors* **20**(15), 4234.

Zuurbier, M., Hoek, G., Hazel, P. v. d. & Brunekreef, B. (2009), ‘Minute ventilation of cyclists, car and bus passengers: an experimental study’, *Environmental health* **8**, 1–10.