**NON-CONTACT BREATH CYCLE ANALYSIS USING RGB-D IMAGING FOR ACTIVITY MONITORING APPLICATIONS**

**Abstract**

Awareness over vital signs monitoring has increased over the past decade which has led to a rise in commercial wearable devices which monitor a person’s health around the clock. However, these devices fail to accurately interpret vital signs under normal physiological activities such as walking, running and requires to be in physical contact with the user. Depth imaging has gained traction since being introduced and can be used for a variety of applications such as Yoga, CT Imaging, Driver monitoring system. Many studies have been carried out to extract vital signs using depth imaging technology. This paper explores the extraction of every respiration cycle of a person along with the inhalation and exhalation time in real-time using depth, RGB imaging under minimal movement. The proposed method makes use of the abdominal-thoracic region as a region of interest (ROI) and relates the small distance change in the chest region with breath and obtains a time-series signal of the user’s breath. A dataset involving 30 people was collected along with the breath ground truth and the breath data from each person was analyzed and reported. Different breathing modalities are evaluated using the dataset and the findings have been discussed. Results show that the quantification of real time breath cycle parameters highly correlates with the ground truth.

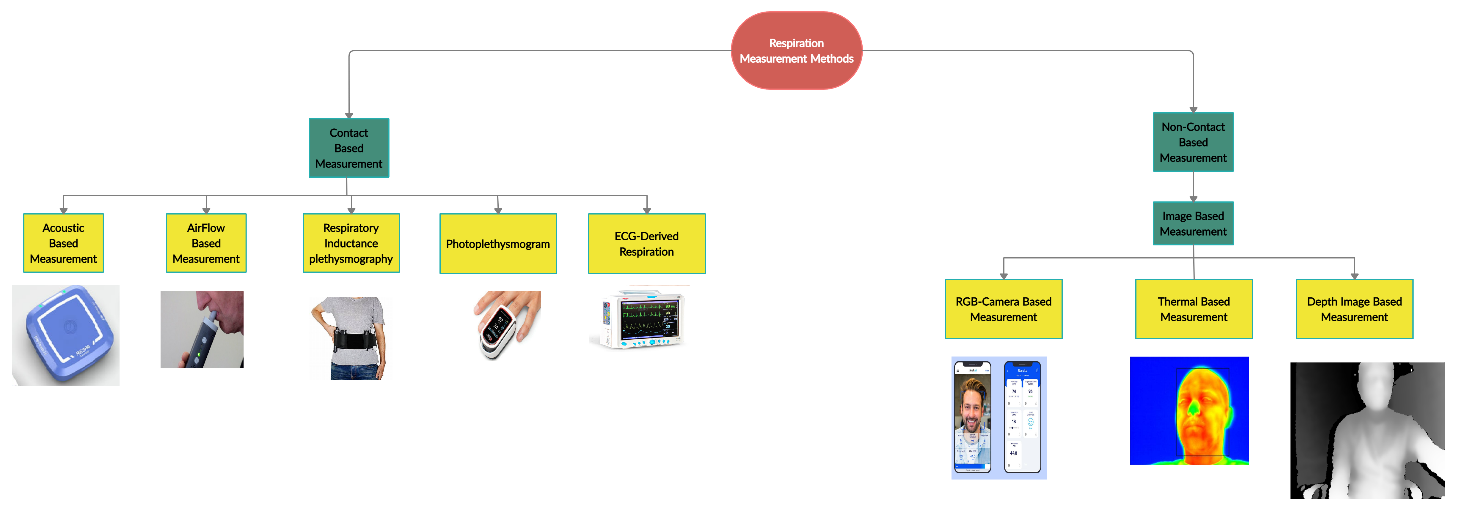
Keyword: Bio-feedback, Breath training, Depth camera, Non-contact, Breath cycle analysis, Skeleton Tracking

1. **INTRODUCTION**

Though respiration is one of the most important vital signs, it has been less utilized for characterizing a person’s health. It can be used in analyzing different pathological conditions such as adverse cardiac events, pneumonia, sleep apnea diagnosis and stress-based conditions such as emotional stress, physical effort, fatigue. Applications such as Yoga, CT Imaging [1], Driver monitoring where a time domain analysis of respiration would provide more informative data has gained traction in the recent years. Most existing methods for respiration estimation discuss a frequency domain-based approach wherein the respiration rate is determined. However, in spite of its importance, very less contributions have been made to study the time domain approach where parameters from every cycle of breath can be used in assessing a person’s health [2][3][4].

Some of the parameters that are least explored and extracted from respiration are inhalation and exhalation time, respiratory resistance, Air flow volume [9]. Of these, the inhalation and exhalation time are represented in the time domain and could be useful in the detection of Respiratory Sinus Arrhythmia (RSA) [8]. Studies have also proven that these parameters modulate the heart rate [6][7]. Thus, detecting the inhalation and exhalation time in real time would not only help in analyzing its effects but also to develop applications that would enable a user to control it. BreathCoach [10] is one such application which uses sensors on a smartphone-based VR and smartwatch to measure the breathing pattern, inter-beat interval. Based on the measurements, an optimal Breathing pattern is computed and given as feedback visually in a VR game-based training platform.

In addition, techniques for regulating breathing pattern help in stress management, relaxation, psycho-physiological state control [11]. Pranayama is one such technique in yoga where breath control with inhalation, retention, exhalation is performed quickly or slowly. The detection and visualization of inhalation and exhalation time periods while performing this technique would assist in further improvement of the technique in achieving metacognitive awareness, emotion regulation.



**Fig. 1 Respiration Measurement Methods**

Fig. 1. shows the methods for measuring respiration. Non-contact measurement techniques have several advantages over contact measurements. In methods like ECG-derived respiration, patients are confined with cables causing discomfort which influences the measurement. In Airflow based measurement, the user is instructed to breathe through a mouthpiece attached to the device while having the nose pinched off which makes it highly risky for critical patients. Acoustic based measurement methods like Respa [13] have been introduced for mainly for respiration measurement during activity. Photoplethysmogram based measurements are widely used to measure heart rate and are highly sensitive to motion and color interference. Non-contact methods of vital sign measurement have been a huge attraction among researchers paving way for numerous proof of concept implementations. However, at this stage, there is a huge need for improvement in performance and adaptability of this method.

This scope for improvement serves as the motivation for this paper where a non-contact method of measuring the user’s breath for different modalities is implemented. Parameters from breath such as inhalation and exhalation time for every breath cycle are determined in real-time. Ground truth is obtained simultaneously to validate the results and its correlation with the proposed method is computed.

A vision-based system using RGB imaging has been explored widely for breath rate and heart rate extraction. The change in skin color due to blood flow is recorded with an RGB camera and used to measure the breath rate in []. A 2D image subtraction technique to detect the movement of chest and abdomen caused by breathing has been reported in []. Similarly, changes in the intensity of light reflected back from the upper chest area is used to analyze breath-by-breath respiratory pattern in []. The lack of accuracy while tracking movement and the necessity of illumination are the major disadvantages of using an RGB camera. This issue has been addressed by placing markers on the user to track movement but in the process becomes a contact based measurement technique.

Several other non-contact measurement methods of breath have also been implemented such as Ultrasound imaging [], thermal sensors [], Radiography and Fluoroscopy imaging sensors []. Other than these methods, depth cameras have been explored a lot in non-contact measurement of respiration. Depth cameras such as Kinect v2, Intel RealSense have been utilized to extract respiration from the user. Depth sensors determine the distance from the camera in the specified Field of View (FOV). This can be visualized and can be used based on the application. Based on the working principle, depth cameras can be classified into Structured and Coded light, stereo depth, Time-of-Flight (TOF).

Structured light cameras project a pre-defined pattern of light onto the scene and senses the change in light pattern to determine depth information. This method was used in [] where few dots were projected on the user’s chest and were tracked using the Structured light camera’s IR sensor. Time-of-Flight measures depth based on the time difference between the light emitted and its return after being reflected by any object. Based on the time difference, the distance between the camera and the user is determined and given as depth data. [] uses TOF based Microsoft Kinect sensor to determine breath and heart rate by mapping chest movements during different kinds of breathing. The user’s mouth was selected as the region of interest (ROI) in the depth images and the corresponding breath was obtained. Heart rate estimation was carried out from the infrared video. Similarly, [] uses frame subtraction to recognize movement in the chest between the frames to determine the respiratory rate using MS Kinect. This has been used in detecting central apnea for people with compromised respiratory reflex.

Stereo vision based depth images rely on simultaneous video acquisition and disparity calculation from two cameras separated by a small distance. The correlation between the acquired images are used to develop the depth images. It also makes use of a laser projector for better accuracy. Also, each breath cycle is analyzed for inhalation and exhalation time detection which can be used in breath-based activities such as Yoga, driver monitoring etc. The inhale, exhale ratio was also calculated which stresses the importance of a balanced breath. The obtained results are compared with a ground truth measurement of breath and validated.

Real-time inhalation and exhalation time detection can be utilized to develop training strategies for controlled breathing as shown in Fig. . Synchronizing breathing in and out at regular intervals with heart beat tends to improve the wellness of the user.

Normal Breathing-10s

Inhalation-4s

Exhalation-2s

Normal Breathing-10s

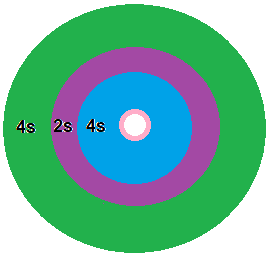
Resting-4s

n=1

n=n+1

n<=5

n>5



**Fig.**

A visual feedback-based breath training could be utilized as a form of therapy for chronic, asthma related diseases.

**2. Related Work**

**2.1 Products for Vital signs measurement**

Commercial applications which obtain breath data with minimal contact with the user are available in the market these days. However, these devices tend to be unreliable during physical activity. An exclusive wearable device which measures breath during activity has been developed by Zansors which uses sound from every breath cycle to compute respiration. It also tracks the ventilatory threshold among other features.

A real-time vital sign measurement using a smartphone camera has been developed by Binah AI. The product uses real time RGB images to track the face continuously and to determine the PPG signal based on intensity changes from which signals such as Respiration, SpO2, Heart-Rate variation are obtained.

Another application based on non-contact vital signs measurement is BreathCoach, a smart unobtrusive system that provides RSA-Breath training using sensors on a smartwatch and smartphone based Virtual Reality game. It measures bio-signals such as Breathing pattern, inter-breath pattern, amplitude of RSA and provides optimal breath pattern based on current and past measurements. The recommended breath pattern is provided as a feedback through VR for effective guidance.

**2.2 Remote Respiration Estimation**

Several methods employing frame-subtraction, intensity calculation and optical flow-based techniques have been explored in literature so far. Depth-based breath acquisition falls under reflective non-contact measurement and has been exploited using devices like MS Kinect, Intel RealSense, Orbbec Astra. Most of the explored methods are indoor-based, short range and normally tested outside a clinical setup. The number of works based on Stereo Vision sensors for respiration assessment is relatively limited since they were introduced recently.

Most of the depth-based respiration measurements rely on the change in motion of the chest or abdomen. It is known that the inhalation and exhalation is due to the expansion and contraction of the user’s chest which can be observed as a change in distance using depth sensors. In most cases, the person of interest would be restricted to small movements. A majority of the work computes respiration by averaging pixels in the region-of-interest and also at higher sampling rates, smaller distance between user and camera.

Al-Naji and Chahl have developed a system for observing abnormal cardiopulmonary events in different sleeping postures under varying illumination. They also determine the respiration rate from the abdominal-thoracic region which will be used in sleep apnea detection. Frame subtraction was used to obtain breath from subsequent frames. Frame contrast and noise were optimized by binarization processing methods such as contrast-limited adaptive histogram equalization and Morphological filtering. The difference between consecutive frames along with the frame rate from the sensor provides a vector of respiratory signals from which the respiratory rate can be determined.

M. Schatz, A. Prochazka et. al. presented a non-contact method of breath cycle detection using MS Kinect along with a comparison study of breath during wakefulness and sleep. The user’s state was classified based on the breath cycles obtained from the depth sensor and not the breath rate since the sampling rate of the device used was not constant. A Bayes Classification method was used to classify the user’s wakefulness and sleeping condition.

F. Benetazzi, A. Freddi. et. al. formulated an algorithm to derive the respiratory rate by measuring the morphological changes in the chest wall. RGB-D cameras were used to identify the human chest and mean depth in the region interest expressed as a time series represented the breath of the user. In addition, inhalation and exhalation time for each breath cycle was determined from the local maxima and minima. The method was validated with a spirometer and was found to have a correlation coefficient of r=0.9292.

Polysomnography based breath signals were explored by M. Schatz, A. Prochazka et. al who also proved that the breathing signals from depth sensors have the same sensitivity to breathing changes as in PSG records. Sleep apnea events were classified with the same accuracy as obtained from the PSG data. Using the Skeletal joints from the Kinect Library, the mean values of difference between two consecutive frames was evaluated over the selected region of interest. A band pass FIR filter was used to eliminate noise along with a second order Savitzky-Golay filter. The filtered signal was divided into smaller overlapping pieces and given as input to the classifier. The classifier marks the occurrence of sleep apnea events over the entire sleep cycle which was validated by a sleep specialist.

M. Valis, O. Vysata et. al. explained the method of using non-contact breath and heart rate to detect medical and neurological disorders. The user’s mouth was chosen as the region of interest from the infrared data while the chest motion was determined from the depth sensor data. Respiratory rates from the Depth sensor, RGB and infrared sensor were compared. Deep and shallow breathing modalities were tested.

J. Kempfle and K. Van Laerhoven have developed a novel approach to monitor a user’s respiration which does not require the user to be in a predetermined position. An occlusion mask and the predicted torso from the previous frame are used to recover the lost pixels in the occluded region. To find the occlusion mask, each pixel in the difference image between frames are compared to a certain threshold that gives a minimum distance to the user’s torso. Taking the mean of the selected torso region is a simple but effective method of extracting the breath signal. The methodology was tested for sitting and standing positions and validated with a respiration belt.

Several methods employing low-cost thermal imaging techniques have also been employed. One such method has been presented by A. Kwasniewska, J. Ruminski and M. Szankin which computed the respiration rate from extremely low-resolution thermal image sequences. A deep learning model to super-resolve images from which RR is estimated has been developed. The results were compared with RR obtained sequences with magnified color changes using the Eulerian Video Magnification algorithm.

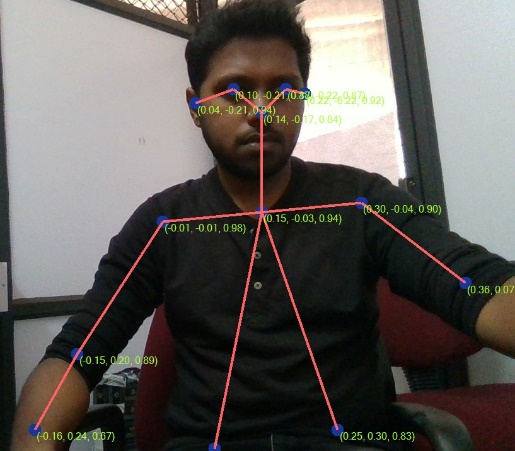
Breath and other vital signs estimated using non-contact methods have a Driver monitoring based on breath data can be used as an indicator for drowsiness. Emotion classification and several other physiological features can be obtained from breath data.

Our method marks the foundation for a standard non-contact real time breath measurement while sitting at a distance within 4 meters. Additionally, each breath cycle is analyzed and the consecutive inhalation and exhalation time are determined. This paper also explores the possibility of using depth sensors in a low illumination environment to obtain breath along with a cycle-by-cycle analysis.

**Data Acquisition**

Stereo vision based Intel RealSense is used to acquire RGB-D images simultaneously with a frame rate of 30 fps at a resolution of 640x480. The camera was placed at a distance of 2-3 m from the user. It uses a rolling shutter technology wherein the frame is captured by scanning the scene horizontally or vertically. It has a high resolution depth, with a narrow field of vision. It consists of two infrared cameras, a RGB camera along with a laser projector. The laser projector is used to improve the efficiency of the depth image. Since, the laser needs to be pointed at the user for a prolonged time, its characterization was tested where the laser power at different distances was checked with a detector. The laser was set at 150 mW throughout the data collection and its power was verified to be in the range of micro-Watts at a distance of 1 foot.

CUBEMOS Skeleton tracking is used along with RealSense for tracking the user’s skeleton joints. It uses depth camera for robust real time results. The cameras differentiate a human from a background, and then identify the position of a number of features or joints, such as shoulders, knees, elbows and hands.  Once those joints are identified, the software connects them into a humanoid skeleton and tracks their position real time. This data can then be used to drive interactive displays, games, VR or AR experiences. Using a total of 18 joints detected, the user’s shoulder and hip joints are used to define the region of interest for breath detection. MATLAB is used for filter design while Python is used for real-time breath monitoring and analysis.



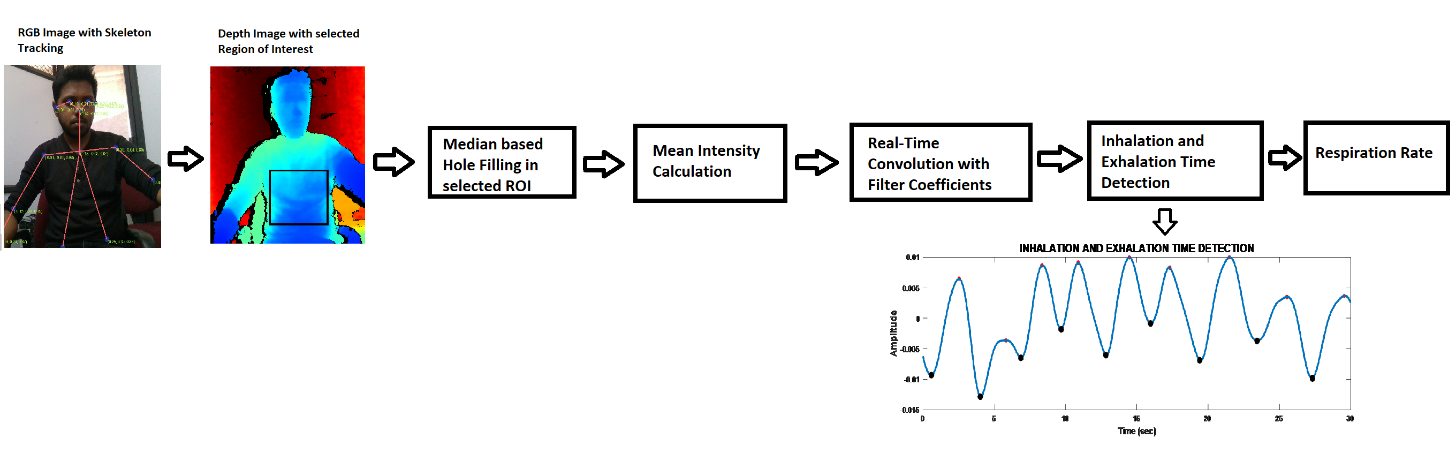
Vernier Go-Direct Respiration belt is used to validate the results from Intel RealSense. It uses a force sensor to determine chest expansion from which breath can be determined. This belt could be tied around the user’s chest and used to transmit data either via USB or through Bluetooth. The breath data sampled at 10 Hz is used simultaneously with the proposed method. Python was used to record the RGB-D videos and also was used in real-time algorithm development. In addition, MATLAB was used in designing the filter coefficients.

**Dataset**

The proposed method was tested with 15 adults for different breathing patterns such as Normal, Slow, Fast breathing. In addition, a period wherein the user holds his breath after a breath cycle was also achieved. Each participant was informed of the research objective and assured of their safety. The data was recorded indoor with the users sitting in front of the camera at a distance of 2-3 meters in a well-lit environment. The users wore the respiration belt and faced the camera for a period of 50 seconds for the mentioned breathing patterns. They were instructed to minimize the movement of their arms during the test since the tightening of the muscles introduces motion artifacts. Both the respiration belt and Intel RealSense were operated simultaneously to obtain data. Finally, data from the respiration belt and Intel RealSense based RGB-D video was recorded and stored.

**METHODOLOGY**

The working of the proposed system to extract breath from depth data captured by Intel RealSense D415 is illustrated in Fig. .



**Fig**

The sequence of RGB images obtained are used as input to the Skeleton Tracking API. The Skeleton Tracking software takes a RGB image as an input and returns the co-ordinates of 18 joints along with their confidence scores. Since the chest wall movement is directly proportional to the breathing pattern in most cases, the region of interest is selected as shown in Fig. . Since depth data has susceptible to noise, holes appear in the depth image which tend to introduce errors. A median-based approach to fill the holes is used where the median value of the selected ROI is mapped to the pixels with no depth data. Once the ROI is selected, an intensity-based approach is used to extract respiration. The selected region R is represented by pixels with R rows and C columns. The mean intensity for a signal of length is given as



The Mean intensity is calculated over a period of time representing the respiration of the user. Several methods have been used to eliminate noise from the calculated mean intensity. Moving Average based methods tend to eliminate the higher frequencies is one of the approaches used to filter noise. However, when used to extract different breathing patterns such as slow, normal, fast breathing, it fails to provide an accurate measure of the user’s breath. Savitzky-Golay filter has been employed in several implementations to smooth the data. However, similar to Moving average method, it tends to eliminate higher frequencies. Hence, an FIR Bandpass filter of order 16 has been used with cut-off frequencies 0.1-1.5 Hz representing a breathing range of 15 – 120 breaths per minute.

The filter is designed offline in MATLAB using data from different breathing patterns. The obtained filter coefficients  are then exported to Python environment for real-time filtering. The filtered signal  is given as



The mean intensity signal M(n) obtained from the region of interest is stored in a separate moving buffer which convolves the noisy signals with the filter coefficients to provide the breath signal.

Thus, the mean values from N frames form a time series of length L representing the breathing pattern of the user. Once the incoming mean intensities are filtered, the inhalation and exhalation time period identification takes place. This involves peak and valley detection in the incoming signals for accurate classification. The difficulty in implementing a peak, valley detection algorithm lies in distinguishing between artefactual and authentic events. This can be overcome by setting arbitrary thresholds to eliminate spurious events. The thresholds may be applied to amplitude, slope or duration or combination of them. Gradient of the filtered signal at every step is checked for any change in sign of the gradient. Peak or valley occurs at a given point if:

* Change in sign of gradient has occurred (positive to negative for Peak and negative to Positive for Valley.
* Previous extrema were labelled as the opposite of the current label.

**RESULTS**

This paper aims to prove that breath extracted from the depth images have the same accuracy as the breath extracted from the contact type sensors. It also determines the inhalation and exhalation time for every cycle in real-time which could be used to monitor the breathing patterns during activities such as Yoga, sleep monitoring, driver monitoring etc.

The mean intensity of depth image pixels in the abdominal thoracic region obtained for every frame provides a time-series signal which on further processing gives the breath of a person. Filter coefficients were obtained using a 16th order FIR Bandpass filter with a passband of 0.1-1.5 Hz which represents a breathing range of 15 – 120 breaths per minute. Fig. 1 shows the spectral analysis for a normal breathing condition where the first dominant frequency occurs at 0.70 Hz (corresponding to 42 breaths per minute) and the second dominant frequency occurs at 0.23 Hz (corresponding to 14 breaths per minute).

The breath rates along with the inhalation and exhalation time periods obtained from RealSense depth images are compared with the Vernier GDX Respiration belt and evaluated. Different modalities such as slow, fast, controlled breathing are explored using this methodology.



**Fig. : Comparison of Frequency spectrum for Slow Breathing obtained for 30 sec**

The frequency spectrum for slow breathing using the respiration belt and Intel RealSense gives the breath rate between 7-12 bpm as shown in Fig. . To evaluate the performance of Depth based Breath signal, data from 20 different users was collected. The methodology was tested with the participants sitting in a chair at distances from 2-4 meters. Data from each participant was collected simultaneously using the respiration belt and RealSense depth images. The breath rate was evaluated visually and also from the dominant frequency in the frequency spectrum. The breathing modalities were classified based on the breath rates. Breath rates up to 15 bpm were considered as slow breathing while 15-25 bpm is considered to be normal breathing. Breath rates above 25 bpm are classified to be fast. In addition, breath in which the inhaled air is held in for a few seconds before exhalation was also obtained from the participants.

The respiration belt uses a Force sensor which senses the movement of the abdominal thoracic region. Inhalation results in movement of the abdominal thoracic region inwards which applies less force on the sensor whereas Exhalation results in the movement of the abdominal thoracic region outwards which exerts a higher force on the sensor. However, depth images are a measure of the distance between the camera and the user. In this regard, when a user exhales, the abdominal thoracic region moves outwards resulting in a small decrease in the distance between the user and the camera. Similarly, when a user inhales, the abdominal thoracic region moves inwards resulting in a small increase in the distance between the user and the camera. This implies the increase in signal amplitude as inhalation and the decrease in signal amplitude as exhalation for RealSense based depth images. However, the respiration belt visualizes the increase in force amplitude as exhalation and the decrease in force amplitude as inhalation. This effect can be seen in the comparison of the breath data obtained from both methods. Each inhalation, exhalation time determined are compared with the ground truth at every instant and its Root Mean Square Error (RMSE) is determined.



**Fig : Normal Breath obtained from Respiration Belt and RealSense**



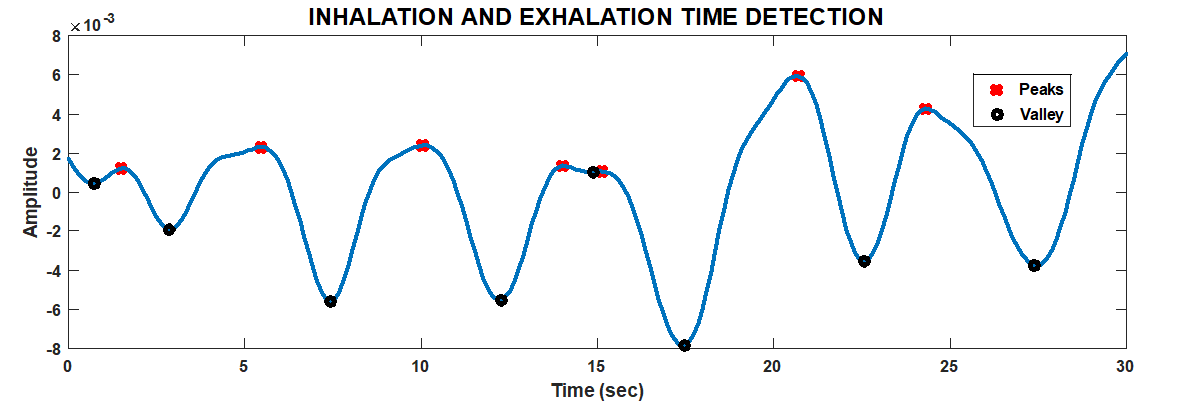
**Fig. Inhalation and Exhalation Time for Normal Breathing**

Fig shows a normal breathing pattern of 18 BPM from the Respiration belt and RealSense based depth images. Every breath cycle is analyzed in real-time and the corresponding inhalation and exhalation time periods are computed in real-time as shown in Fig. . The inhalation and exhalation time periods for both methods are compared for every cycle and its corresponding RMS error is determined. The exhalation time from the belt and the inhalation time from the depth based breath was used to compute the RMS error for inhalation in the depth based method. Similarly, the inhalation time from the belt and the exhalation time from the depth based method was used to compute the RMS error for the depth based method. The RMS error for inhalation is determined to be 0.9014 while for exhalation it was computed as 1.2673.

The inhalation and exhalation time as shown in Fig. can be used to develop a controlled breath based gaming exercise. Regulating the inhalation and exhalation time period forms the core of most of the breathing exercises in yoga. Hence a method of visual feedback which helps a user to control and monitor his/her breathing patterns as shown in Fig. can be achieved with this method. We have further explored different breathing patterns using both methods.



**Fig. Slow Breath obtained from Respiration Belt and RealSense**

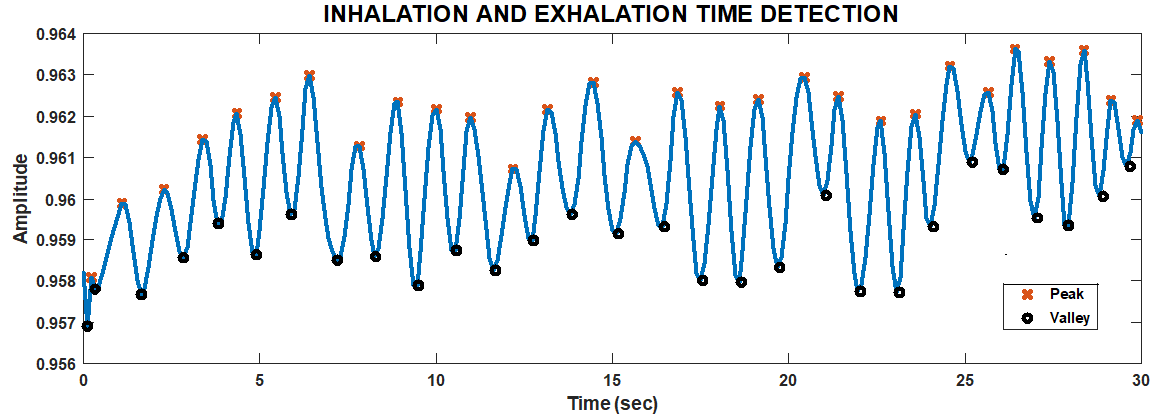


**Fig. Inhalation and Exhalation Time for Slow Breathing**

Fig shows the slow breathing pattern with 14 BPM obtained from both methods and its corresponding inhalation and exhalation time for every cycle. This breathing pattern helps in relaxation and lowers the blood pressure. The inhalation and exhalation time periods for both methods are compared for every cycle and its corresponding RMS error is determined. On comparing the inhalation and exhalation from both methods, the RMS error for inhalation during a slow breathing pattern is computed as 0.5672 while for exhalation it is determined to be 0.9425.



**Fig. Fast Breath obtained from Respiration Belt and RealSense**

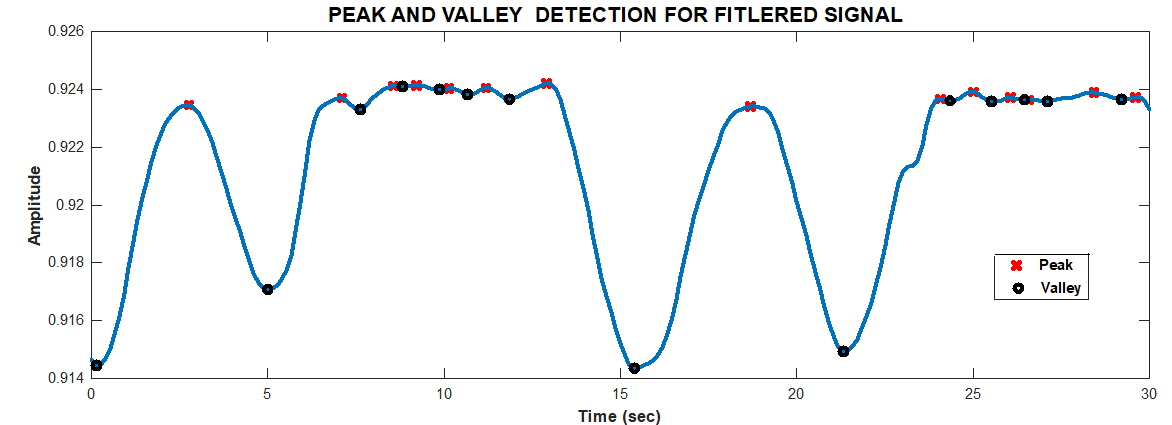


**Fig. Inhalation and Exhalation Time for Fast Breath sequence**

Fig. shows rapid breathing modality data obtained from the respiration belt and RealSense based depth images. Its corresponding inhalation and exhalation time periods are given in Fig. . The breath rate was computed to be 50 BPM. This breathing pattern increases the efficiency of the digestive organs, lungs and heart. The inhalation and exhalation time periods for both methods are compared for every cycle and its corresponding RMSE error is determined. Comparing both methods, the Root Mean Square error for inhalation during a fast breathing sequence is 0.6456 while the same for exhalation is 0.4896.



**Fig. : Slow Breath with a hold period**



**Fig. Inhalation and Exhalation Time for Holding Breath**

Fig. introduces a breathing modality where breath is put on hold for a few seconds and visualized. The corresponding inhalation and exhalation time periods are computed as shown in Fig. . Slow breathing with a hold in between inhalation and exhalation was carried out with the participant moving slightly in the Field of View of the camera and the obtained breath is plotted in Fig. . It can be seen that due to movement of ROI of the participant, high frequency noise is observed along with the breath signal. Similar to the other modalities, the inhalation and exhalation time periods are determined for every breathe cycle. The RMS error for inhalation from both methods is computed to be 5.5289 while for exhalation it is found to be 4.98.



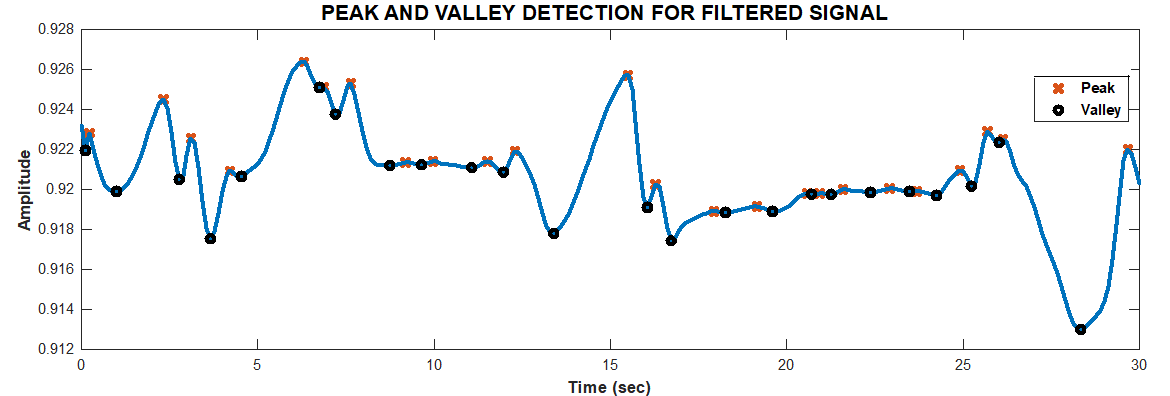


Fig. show the slow breathing modality with a hold between inhalation and exhalation when there is a slight movement from the normal sitting position.

From the above results, it can be seen that the filtered signal obtained for Normal, Fast breathing pattern is smooth when compared to slow and hold/rest breathing pattern. It is challenging to design a filter that is applicable for different kinds of breath pattern using conventional FIR filtering method.