Measurement of imperceptible breathing movements from Kinect Skeleton Data

Filipe Manuel Baptista Marques

Thesis to obtain the Master of Science Degree in Electrical and Computer Engineering

Supervisor: Prof. Dr. Alexandre José Malheiro Bernardino
Dr. Daniel Simões Lopes
Prof. Dr. Joaquim Armando Pires Jorge

Examination Committee
Chairperson: Prof. Dr. João Fernando Cardoso Silva Sequeira
Supervisor: Prof. Dr. Alexandre José Malheiro Bernardino
Members of Committee: Prof. Dr. Ana Luísa Nobre Fred

November 2017
Acknowledgments

I was fortunate to have the assistance and support of several people, who helped me not only in successfully complete this thesis, but also in boosting my education and skills. I would like to acknowledge and thank them in this section.

First of all, I would like to thank to my supervisors, Prof. Alexandre Bernardino, Prof. Joaquim Jorge and Dr. Daniel Simões Lopes.

To the Lisbon Biomechanics laboratory team for all help that I needed, in particular to Prof. Miguel Silva and Dr. Sérgio Gonçalves. Dr. Sérgio, thank you for always been present during the experimental work and for companionship, which facilitated a lot all the procedures.

I also acknowledge the financial support supported by national funds through the Portuguese Foundation for Science and Technology with reference IT-MEDEX PTDC/EEI-SII/6038/2014.

To my mother, who have always encouraged and supported my education. To my sister Filipa, for the always welcome company and distraction from work. To my brother João, for introduced me Electrical and Computer Engineer and help me in the hard times. To all my friends who believed in me, and understand the lack of time due this thesis.

And last, but not the least, to my girlfriend Carolina, which was essential to this work is finished, supporting me in the good times and in hard times, giving me the strength in the moments I needed the most. Thank you for everything Carolina, I love you.

Thank you all!
Abstract

This work presents a signal processing technique to measure respiratory frequency by relying only the skeletal tracking data of an inexpensive depth camera. Our approach aims to be useful for reading the breath frequency automatically in a non-invasive and markerless manner, using the skeletal joints from Kinect data. This study applies a pass-band filter that isolates the breath micro-movements of skeleton's joints to compute the respiratory frequency. To evaluate the system, we collected data from 28 subjects in 4 different acquisition modes. To assess our approach, we considered as ground truth data acquired from the Spirometer device.

The results show that it is possible to extract reliable kinematic data from skeletal data to calculate the respiratory rate. It was found a pattern, using the axis that had better results for each acquisition mode. For both body poses (seated and standing) the best axis was an axis from the left shoulder. The best results for each mode had an error around 2 bpm when compared with the results from spirometer. Analyzing the results of the axis that had the best result in each acquisition the error decreases a lot. This system gains an advantage by achieving results with a minimal processing, which can be useful to apply in low processing devices or to run in the background without causing any negative processing impact. It was possible to identify two limitations in the method proposed in this work. The system only works with static subjects and the thickness or loose clothes can interfere the measurement performance.

Keywords: breathing rate, signal processing, Kinect
Resumo

Este trabalho apresenta uma técnica de processamento de sinal para medir a frequência respiratória usando apenas dados de uma câmera de profundidade de baixo custo. A nossa abordagem pretende ser útil para medir a frequência respiratória automaticamente, de forma não invasiva, usando os pontos do esqueleto criado internamente pelo Kinect. Este estudo pretende aplicar um filtro passa-banda aos pontos do esqueleto de forma a isolar os micro movimentos da respiração e assim calcular a respetiva frequência. Para avaliar o sistema, foram recolhidos dados de 28 indivíduos, em 4 modos de aquisição diferentes, comparados com os dados adquiridos a partir de um espirómetro, considerado como dispositivo de referencia.

Os resultados mostram que é possível extrair dados cinemáticos confiáveis a partir de dados do esqueleto do Kinect de forma a calcular a frequência respiratória. Avaliando os resultados, foi encontrado um padrão: o ponto que apresentou melhores resultados em todos os modos de aquisição foi o ponto do ombro esquerdo. Os melhores resultados para cada modo de aquisição tiveram um erro a rondar os 2 bpm, quando comparados com o resultado do espirómetro. Se for usado o eixo que teve o melhor resultado em cada aquisição, o erro diminui bastante. Este sistema demonstra ser vantajoso ao obter resultados sem ser preciso um elevado processamento dos dados, o que pode ser útil para aplicar em dispositivos de baixo processamento ou para executar em segundo plano sem causar qualquer impacto de processamento. Foi possível identificar duas limitações no método proposto neste trabalho. O sistema só funciona com sujeitos estáticos e a espessura ou o uso de roupas largas podem interferir no desempenho da medição.

Palavras-chave: frequência respiratória, processamento de sinais, Kinect
# Table of Contents

ACKNOWLEDGMENTS ........................................................................................................... I

ABSTRACT .......................................................................................................................... III

RESUMO................................................................................................................................ V

TABLE OF CONTENTS ........................................................................................................ VII

LIST OF FIGURES .............................................................................................................. IX

LIST OF TABLE .................................................................................................................. XI

LIST OF ACRONYMS .......................................................................................................... XII

1. INTRODUCTION............................................................................................................ 1
   1.1. MOTIVATION ........................................................................................................ 2
   1.2. PROBLEM STATEMENT ..................................................................................... 3
   1.3. CONTRIBUTIONS ................................................................................................ 4
   1.4. PUBLICATION .................................................................................................... 4
   1.5. ORGANIZATION OF THE THESIS ..................................................................... 4

2. LITERATURE REVIEW .................................................................................................. 7
   2.1. MAGNIFICATION ALGORITHMS ....................................................................... 8
   2.2. SIMILAR PROTOTYPES ..................................................................................... 11
   2.3. PRISM-BASED METHOD .................................................................................. 14
   2.4. DIFFERENCE OF BREATHING BETWEEN WOMEN AND MEN ..................... 14
   2.5. BREATHING MONITORING MEDICAL APPLICATION ...................................... 15
   2.6. SKELETAL JOINTS ANALYZE .......................................................................... 15

3. THEORETICAL BACKGROUND .................................................................................. 19
   3.1. BREATHING CYCLE .......................................................................................... 20
   3.2. CROSS-CORRELATION ..................................................................................... 23
   3.3. MICROSOFT KINECT (v2) ................................................................................ 24
       3.3.1. Time of Flight .............................................................................................. 24
       3.3.2. Skeletal Tracking ....................................................................................... 25
       3.3.3. Skeletal Joints ........................................................................................... 26
   3.4. SPIROMETER ....................................................................................................... 28
### 4. METHODOLOGY

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.</td>
<td>EXPERIMENTAL PROTOCOL</td>
<td>31</td>
</tr>
<tr>
<td>4.2.</td>
<td>EXPERIMENTAL SETUP</td>
<td>34</td>
</tr>
<tr>
<td>4.2.1.</td>
<td>Microsoft Kinect v2</td>
<td>34</td>
</tr>
<tr>
<td>4.2.2.</td>
<td>Spirometer</td>
<td>35</td>
</tr>
<tr>
<td>4.3.</td>
<td>PASS-BAND FILTER</td>
<td>37</td>
</tr>
<tr>
<td>4.4.</td>
<td>METHODS</td>
<td>38</td>
</tr>
<tr>
<td>4.5.</td>
<td>CROSS-CORRELATION</td>
<td>39</td>
</tr>
<tr>
<td>4.6.</td>
<td>CALCULATION OF BREATHING RATE</td>
<td>41</td>
</tr>
<tr>
<td>4.7.</td>
<td>EVALUATION METRICS</td>
<td>41</td>
</tr>
</tbody>
</table>

### 5. RESULTS

- Page 43

### 6. DISCUSSION

- Page 51

### 7. CONCLUSIONS AND FUTURE WORK

- Page 55

### 8. REFERENCES

- Page 58

**APPENDIX A**

- INFORMED CONSENT | Page 61

**APPENDIX B**

- BIOSIGNALSPLUX EXPERIMENTAL WORK | Page 65
  - BioSignalsPlux | Page 65
  - PRELIMINARY EXPERIMENTAL WORK | Page 66
  - RESULTS OF THE PRELIMINARY EXPERIMENT | Page 67

**APPENDIX C**

- EVM APPROACH | Page 71
List of Figures

Figure 1 An example of using Linear EVM framework for visualizing human pulse.[2] ......................................... 9
Figure 2 - Reliability of displacement measures from the traditional and Microsoft Kinect system ...... 16
Figure 3 Graphic Representation of pressure changes in lungs [26]................................................................. 20
Figure 4 Graphic representation of thorax extending during inspiration and thorax reducing during expiration .......................................................................................................................... 21
Figure 5 Explanation of how synchronism works through cross-correlation. (Top) The signals misaligned; (Center) Result of cross-correlation between the signals above, in interval of lag [-4,4] s; (Bottom) Signals align after correcting their delay.[27] ................................................................. 23
Figure 6 Schematic of how the ToF method works. [31].................................................................................. 24
Figure 7 User face position for it to be detected............................................................................................... 25
Figure 8 Default range of camera....................................................................................................................... 25
Figure 9 Steps taken to compute the skeletal joints. From an input depth image, is inferred the body parts (Colors indicate the most likely part labels at each pixel, and correspond in the joint proposals), and the 3D location of the body joints are estimated.[28] .................................................................................. 26
Figure 10 Average precision of 16 joints from Kinect skeleton ........................................................................ 27
Figure 11 – Materials contained in the Exercise Physiology System of ADInstruments................................. 28
Figure 12 - Block diagram of the methodology................................................................................................. 30
Figure 13 - Schematic of the acquisition procedures. ....................................................................................... 31
Figure 14 - Face mask used in the experimental study. ....................................................................................... 32
Figure 15 - Photo of a participant running on the treadmill. .............................................................................. 33
Figure 16 – Photos of how the acquisitions were made in the laboratory. (left) Seated acquisition; (right) standing acquisition............................................................................................................. 33
Figure 17 - Skeleton positions relative to the human body [29] ....................................................................... 34
Figure 18 – (Left) Depth map from Kinect with the five joints used marked in the image in red; . (Right) 3-D axis orientation representation with the location of the Kinect sensor as reference.[33] ........ 35
Figure 19 – Graphic representation of how the Spirometer device was attached to the remaining devices.[34] .................................................................................................................................. 36
Figure 20 - Example of a signal before and after the filtration. (Left) Unfiltered signal, (Right) Filtered signal ........................................................................................................................................ 37
Figure 21 - Graphic representation of the triangles used in the methods two and three ............................. 38
Figure 22 – Alignment between a y axis of spine shoulder point and the spirometer signal. (Left) Unsynchronized signals; (right) Synchronized signals........................................................................................................ 39

Figure 23 - Alignment between a x axis signal from triangle 1 and the spirometer signal. (Left) Unsynchronized signals; (right) Synchronized signals........................................................................................................ 40

Figure 24 - Alignment between a signal of the variations of area from triangle 3 and the spirometer signal. (Left) Unsynchronized signals; (right) Synchronized signals........................................................................................................ 40

Figure 25 - Histogram of the median values between the axes of method 1 along all acquisitions. .............................................................. 40

Figure 26 - Mean and Standard Deviation in the method of points axes in the four modes of acquisition. (Review joint corresponding to the number of the horizontal axis in Figure 18-left).......................... 44

Figure 27 - Mean and Standard Deviation in the method of Barycenter axes in the four modes of acquisition. (Review triangle corresponding to the number of the horizontal axis in Figure 21). ............................ 45

Figure 28 - Mean and Standard Deviation in the method of the triangle area in the four modes of acquisition. (Review triangle corresponding to the number of the horizontal axis in Figure 21). ............................ 45

Figure 29 - Heat map of the error along all tested axes. ................................................................................................................................. 46

Figure 30 - Graphics of the breathing rates expected from Kinect joints and from Spirometer. ............................... 47

Figure 31 – Results of the best axis for each acquisition for each mode................................................................. 48

Figure 32 – Comparison of the percentage of error between method 1 without outliers and the results using the best axis for each acquisition. ........................................................................................................ 49

Figure 33 – (Left) Wireless 8 channel hub; (Right) Piezoelectric respiration sensor. ........................................ 65

Figure 34 - Mean and Standard Deviation in the method of points axes in the four modes of acquisition using the BioSignalsPlux. (Top left) Seated acquisition before exercise; (Top right) Standing acquisition before exercise; (Bottom left) Seated acquisition after exercise; (Bottom right)........ 67

Figure 35 – (Left) Example of a plots of a bad acquisition, before and after filtering the signal; (Right) Example of a plots of a good acquisition, before and after filtering the signal............................. 69

Figure 36 - Example of the variations of a signal from skeletal joints after filtering and magnifying it (in 30 s).................................................................................................................................................. 71
List of Table

Table 1 - Results of the best axis for each acquisition for each mode............................................. 48
Table 2 - Error of the best axis in each acquisition .............................................................................. 68
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D</td>
<td>One-Dimensional</td>
</tr>
<tr>
<td>3-D</td>
<td>Three-Dimensional</td>
</tr>
<tr>
<td>BPM</td>
<td>Breaths Per Minute</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiography</td>
</tr>
<tr>
<td>EVM</td>
<td>Eulerian Video Magnification</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of View</td>
</tr>
<tr>
<td>FPS</td>
<td>Frames Per Second</td>
</tr>
<tr>
<td>HR</td>
<td>Heart Rate</td>
</tr>
<tr>
<td>HRV</td>
<td>Heart Rate Variability</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent Component Analysis</td>
</tr>
<tr>
<td>INESC-ID</td>
<td>Instituto de Engenharia de Sistemas e Computadores, Investigação e</td>
</tr>
<tr>
<td></td>
<td>Desenvolvimento</td>
</tr>
<tr>
<td>MIT</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>PC</td>
<td>Personal Computer</td>
</tr>
<tr>
<td>RGB</td>
<td>Red-Green-Blue (color model based on additive color primaries)</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>RR</td>
<td>Respiration Rate</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SDK</td>
<td>Software Development Kit</td>
</tr>
<tr>
<td>ST</td>
<td>Skeletal Tracking</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>ToF</td>
<td>Time-of-flight</td>
</tr>
<tr>
<td>VPM</td>
<td>Vicon Peak Motus</td>
</tr>
<tr>
<td>WAM</td>
<td>Weight Arithmetic Mean</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
1.1. Motivation

Monitoring a patient's breathing cycle is very important to identify early signs of serious complications that can be manifested by: decreased oxygen saturation; respiratory acidosis; mental agitation; cardiac disturbances including cardiac arrest. Monitoring the breathing cycle can be an important mean to detect the increase of respiratory rate that can be a consequence of hypercapnia, hypoxia or metabolic acidosis. [1]

Therefore, measuring the respiratory rate may support the early identification of high-risk patients. A published analysis of more than a million sets of vital data of patients in a United States hospital found an association between the deviations of the respiratory rate from normal. Thus, it is important to track the breathing rates to prevent major risks, the abnormal respiratory rates should be carefully monitored.[1]

Recently, Massachusetts Institute of Technology (MIT) researchers have developed a revolutionary system that allows magnifying subtle changes, not visible to the naked eye in a video [2]. This system has already been applied in various medical applications, such as detecting the heartbeat and calculating the respiratory frequency, through video sequences that are magnified (i.e. Ming-Zher Poh 2010 [3]; Guha Balakrishnan 2013 [4]; Chambino 2013 [5]). The first aim for this study is to develop a technique similar to the one used by the MIT researchers, to magnify subtle motions invisible to the naked eye that could be used to achieve biomedical data as breathing rate. In contrast to these approaches, that filter and magnify video sequences, in this work it was filtered one-dimensional (1-D) signals from skeletal joints from Microsoft Kinect. This approach aims to be more efficient to run in real-time, once the level of data processing is less complex.

The Microsoft Kinect is a low-cost motion sensor that contains an Red-Green-Blue (RGB) camera, an infrared sensor, a microphone and it also has an embedded processor. This sensor can do an Interactive human body tracking and this has been used in various applications including gaming, human-computer interaction, security, telepresence and health-care[6]. With the introduction of real-time depth cameras, those tasks have been simplified, and various researchers are studying new possibilities for this system.

The capabilities of Kinect sensor oriented to medical applications can be a great asset to quickly analyze some vital signs that can be reported to the hospital of the area or, at least, may support the early identification of high-risk patients by raising an alarm to inform the subject about the variations of vital signs. As can be seen in the article of Hossein Mousavi Hondori and Maryam Khademi in 2014 [7] the number of annual publications by Pubmed of using Microsoft Kinect on therapy and rehabilitation are growing since 2010, that evidence the great potentiality of this sensor.

These applications can also be an interesting input for gaming, so the game can deal with the emotions of the subject and the gaming companies can understand how the clients react to some tasks. This knowledge can also adapt the game to the difficulties to each gamer.
At the moment, the skeletal joints from Kinect are popularly used for games, but can be used in a wider diversity of applications [8]–[11]. In this work, it was studied the reliability of this sensor to measure breathing rate, using a custom designed filtering method applied to the raw skeletal data.

1.2. Problem statement

The detection of physiological information from systems without the use of wires or devices in contact with the body of the subject is something that has been investigated for several years by researchers. The benefits of these type of systems are multiple specially for medical purposes. For example, in a waiting room people get to stay a few hours without any support, so a remote analysis of vital signs can detect in time any changes in the patient's clinical picture and alert the responsible health professionals in time. These types of systems can also be a good asset, for example, in the gaming industry where physiological information can be used to make the games more interactive and react to the player's movements.

To address the remote measurement of breathing rate the scientific community has developed prototypes that use the depth images, essentially from the Kinect sensor due to the low cost of these devices. The main approach is to detect the volume variations in the chest area of the subjects [8], [12]–[17]. A common issue of these approaches is the high computational cost of the solutions, since those approaches process high amount of data, to extract the breathing rate.

In this thesis, it was made a novel approach. It was used directly the skeletal joints from Microsoft Kinect to detect the breathing rate. The Kinect provides five joints in the torso area that can be used for this purpose which will be used on this work.

The work can be divided into three main goals. The first objective is to determine if the skeletal measurements from Kinect data have any physiological information. The second objective is to study if it is possible to measure the breathing cycle from this data. The third and last objective is to analyze the reliability of these signals, validated with a more accurate device.

The solution to this problem should be computationally efficient. In other words, should be possible to be computed in real-time or at least with low time delay [18]. It is important to ensure results with low processing, to allows a low processing device to run this system.

The signals from skeletal joints were filtered to make observable the imperceptible movements of a person during the respiratory cycle. It was necessary to measure breathing rate and compare it with a ground truth device. This measurement system was tested with 28 subjects and compared with a Spirometer, that is not a certified medical device but as an accurate system to measure the air flow. There were made different types of acquisitions, seated and standing, after exercise and in rest. From all tested signals extracted from Microsoft Kinect it is important to assess if it is possible to measure the breathing rate, and what is the most accurate signal.
1.3. Contributions

The contributions of this thesis are of various natures: scientific content has been produced and reported in an international conference, and technical content was materialized into one prototypes of an application to the skeleton joints from Kinect v2.

These contributions are categorized in the following bullet points:

- Literature Review;
- Give a new approach for the current problem;
- Test the approach with a relevant number of people with a ground truth device;
- Comparison with related work

1.4. Publication


1.5. Organization of the thesis

This work presents a signal processing technique to measure respiratory frequency by relying only on the skeletal tracking data of an inexpensive depth camera. Our approach aims to be useful for reading the breath frequency automatically in a non-invasive and markerless manner. This study applies a passband filter that improves the signal to noise-ratio of the micro-movements of skeleton's joints that are related to the respiratory frequency.

In Chapter 2 are reviewed articles about the Eulerian Video Magnification algorithm and its medical applications, including in section 2.1 similar prototypes of breathing rate measurement and other experimental works.

Chapter 3 gathers fundamental concepts used in this work. It starts in section 3.1 by explaining the human breathing cycle and how the breathing affects body fluctuations. Then, in section 3.2 is explained how to align the signals acquired from the Kinect and the Spirometer in order to compare both methods. In sections 3.3. and 3.4 is explained the experimental setup of this work, namely the Microsoft Kinect v2 and Spirometer.

The Methodology of this work is described in Chapter 4. In section 4.1 is presented the Experimental Protocol, which describes how the experimental work was conducted with the volunteers. In section 4.2 is presented the Experimental Setup, which explain the implementation aspects in detail. In the following sections is explained the filter used to extract the breathing movement from the skeletal joints signals,
the methods used to detect the breathing movements through the skeletal joints, how the signals are synchronized, how it was computed the breathing rate and lastly how the system was evaluated.

The results of the experimental work were presented in the Chapter 5 and the discussed of them in the Chapter 6.

Finally, in Chapter 7 it is summarized the document and presents the main conclusions of the work performed. Additionally, a list of several suggestions for future research work to be built on top of the work herein presented is described.
Chapter 2

Literature Review
2.1. Magnification Algorithms

This work, from the beginning, had as main objective to magnify subtle movements of a subject and measure biometric data, such as respiratory rate or cardiac rate. It was first tested an algorithm created by researchers from MIT, which is called Eulerian Video Magnification (EVM) that magnify and filters micro movements and makes them visible. This method was first initialized by Liu [19] that uses a Lagrangian approach. Then, in 2012, Hao-Yu [2] presented the Eulerian Video Magnification that was used as a starting point for this work.

Liu et. al. [19] proposed, in 2005, a motion magnification technique, based on video footage manipulation. The goal was to amplify subtle motions invisible to the naked eye present in apparently static scenes. The motivation was that making these motions visible and available for further analysis could reveal innumerable interesting behaviors hidden in different objects, environments, and subjects. This technique is recognized in later related works as a Lagrangian approach. This approach relates to reliably estimation motions. For this, it does the clustering of pixels whose motions should be magnified together. The feature points of the image are analyzed to group pixels according to his local intensity configurations, once that are all promising candidates for finding reliable motion trajectories. The major limitation of this approach is the artifacts that appear between unamplified and amplified pixels.

Hao-Yu Wu et. al. [2] presented in 2012 a different approach to the one presented above in order to magnify variations in video sequences. As opposed to the referred Lagrangian approach, they were inspired by a Eulerian perspective, and is known as EVM. Instead of explicitly tracking pixel trajectories, they analyze the variations of each pixel over time and magnify each pixel's variation. Relying on different filters suitable for the intended application, in that work, the authors were able to magnify pixel variation both in temporal changes as well as in position changes. Several demos were provided, such as sequences where it is possible to measure pulse rate due to redness variations in the face of a person being filmed, as it can be seen in Figure 1, and the tracking of breathing rate by magnifying around chest movements of a sleeping person. These results were also compared to the ground truth in order to verify the accuracy and set forth great results. This technique proved to surpass the Lagrangian approach in some points such as having the possibility of magnifying color and motion using a single footage. Additionally, this technique is less computationally expensive and does not require further in-painting of hidden parts in between frames. Comparing both approaches, the authors admitted the Lagrangian approach would perform slightly better when using large magnification factors and specifically enhance motions in fine point features. Limitations to the Eulerian technique relate to its noise sensitivity due to the fact that, most times, the noise present in the footage has a much higher amplitude than the variations to be magnified.
EVM Medical Applications

The amplification of small invisible motions and changes in videos presents a whole new range of applications both in medical and non-medical fields. As proposed by the authors of EVM, quite useful applications relate to the measurement of vital signals, using color amplification, analyzing the subjects face, with great results in subjects of different complexions and even wearing makeup. The motions of the blood vessels could also provide a measure for heart-rate when it comes to the magnification of spatiotemporal motion itself (as opposed to solely chromatic variation amplification).

Most applications concern in health and in/out patient monitoring, especially when using real-time processing, technique, which is both inexpensive and effective [2]–[5]. In the following paragraphs will be explained three medical applications that used the EVM algorithm to extract the desired signal.

Ming-Zher Poh [3] proposed in 2010 measuring multiple physiological parameters, as respiratory rate, heart rate and heart rate variability (HRV), using a simple webcam via analysis of the subtle color changes in the skin caused by blood circulation. This method relies on blind source separation by independent component analysis. The results were compared to those obtained by Food and Drug Administration (that is a federal agency of the United States Department of Health and Human Services) approved sensors. The method is motion tolerant and it is shown how it can be extrapolated to several subjects simultaneously. They achieved a strong correlation between the results from standard sensors and the ones measured. The results achieved were considered taking into account several limitations, such as those imposed by the fluctuating sampling rate of the web-camera and the use of a standard PC image acquisition. Also, the sampling rate was much lower than the ones recommended for HRV analysis but, by interpolation, high correlations were reached.

Guha Balakrishnan et. al. [4] in 2013 also proposed measuring the pulse in subjects but his approach was based in motion magnification, amplifying only the small movements in the head related to cardiac activity, by separating these from the range of other involuntary motions on the head-neck section of the
body. To evaluate this method, it was tested in eighteen subjects for 70-90 s. It was compared the results from this method and the Electrocardiography (ECG) monitor that was used as ground truth. The results show that good results with a mean error of 1.5 % of the heartrate. The major limitation is that this only works with static subjects.

Pedro Chambino [5] proposed, in 2013, to implement the EVM method on smartphones to monitoring heartbeat in real-time. For this purpose, first it reads frames from the smartphone’s webcam. For each taken frame, it detects faces and matches them with previously detected faces to track multiple faces. With the collected information, the following steps must be performed: EVM (to magnify detected face's rectangle); Signal noise (verify the quality of the signal - if it is noisy that face’s signal is reset and marked as invalid); Detrending & Normalization (if the current face’’s signal is not too noisy, then detrend and normalize the signal in order to facilitate further operations); Validate signal (verifying its shape and timing, if the signal is given as invalid, it is kept as valid for some time because the validation algorithm may miss some peaks); Calculate beats per minute (if the current face’s signal is valid, it is then used to estimate the person’s heart rate). Lastly, Processed frame, the resulting frame with each magnified face rectangle added back to the original frame. In comparison with measurements of ViTrox application and to the readings of a sphygmomanometer it is concluded that the implemented algorithm has worse estimations with higher heart rates (>70 beats per minute). The biggest limitation is that the algorithm is not efficient enough to run on the smartphone in real-time.

The EVM algorithm promote the path to detect motions that are invisible in the naked eye. This method was used in several medical applications in order to get important information as the vital signs without intrusive devices that are currently used in the hospital and clinics. This type of wireless prototypes may be used in future due to them accurate results in ideal conditions. This thesis work used this algorithm as starting point, but it was concluded that the spatial filter is secondary once it was used 1-D data, in contrast with these systems that used video sequences. Therefore, it was used this algorithm as motivation and an accessible MATLAB function that contains a pass-band filter, that was used to filter the signal temporarily [2].
2.2. Similar Prototypes

In this section, similar prototypes that use depth cameras to obtain biometric information, from a subject, will be described, especially the methodologies used, and the results obtained.

Meng-Chieh Yu et al. [12] proposed, in 2012, a method to measure human chest wall motion for respiratory volume estimation without any physical contact in the subject. This method is based on depth image sensing technique from Microsoft Kinect sensor. In this system, the user needs to sit in a chair in front of a depth camera. The Region of Interest (ROI) are three different areas, left thorax, right thorax, and abdominal area. After the chest wall mask is determined to define the ROI, the respiratory volume could be calculated. The volume changes are obtained by calculating the difference of the current depth image and the reference depth image (the initial). A spirometer was used as ground truth in this study to validate the current method. The experimental work was performed with twelve healthy participants, each one performed sixteen acquisitions (which includes shallow, middle, and deep breathing). The results show that this method is reliable while the morphology of the chest wall changes, with r=0.966 (p<0.001). In addition, it was established that the correlation coefficient of the respiratory volume in the whole thorax and abdomen change in opposite phases. The results also shown that the amplitude of the volume signal from the right side of thorax is substantially larger than the volume signal from the left side of the thorax. There were two major limitations in this study. First, the rigid forward movements of the user affect the measure accuracy. Second, if the subject were thick or loose clothes, it is hard to measure the respiratory volume well.

Narhan Burba et al. [13] proposed, in 2012, two approaches for unobtrusively sensing subtle movements. The first approach to compute the respiratory rate through measuring the fluctuations of the subject’s chest. And the second to detect specific type of fidgeting behavior, i.e. “leg jiggling”, by measuring high-frequency vertical oscillations of the user’s knees. For the first approach, that is the one that is related to this work, the first step was to isolate the chest area in the depth map, using the skeletal joints from Kinect, this ROI only change if the skeletal joints move to a distance larger than 10 pixels. Then, was searched for cyclical fluctuations in this area. To measure the breathing rate, it was computed the frequency of the cyclical patterns. The principal limitation is the movement of the subject if there is any movement it momentarily disrupts the breathing measurement until a stationary pose is resumed.

Natascia Bernacchia et al. [14] proposed, in 2014, a measurement method using only a Kinect sensor to calculate the heart rate (HR) and respiratory rate (RR). The experimental setup used for this study consists of a 4 leads ECG device, a spirometer device, a data acquisition system, and a laptop. The data acquired from the Kinect is acquired synchronously with the ECG and spirometer on a laptop and stored in the same device. The Kinect sensor analyzed three different areas: the neck area, thorax area, and the abdominal area. The first two are used to extract the cardiac activity and the last one to extract the respiratory activity. The experimental work was done with 10 healthy subjects (5 males and 5 females), that perform the acquisitions in supine position on a rigid bed. Each acquisition took 40 seconds. After all acquisitions were taken, the three regions of interest were manually selected, the means of the depth distances in ROI were computed frame by frame and used as input to the
Independent Component Analysis (ICA) that return two output signals (heart and respiratory activity computed signal). The heart rate values were computed with the main peak of the power spectrum density. The respiratory rate values were computed with one of the output signals of ICA algorithm using a wavelet decomposition. The results show a bias deviation of 10 ms for HR measurement and no bias for RR measurement, while an uncertainty (K=1) of 6 % and 9.7 % is reported for HR and RR respectively. The principal limitation is the number of subjects used in this study (only 10), and it should be tested the subjects in different postures (sitting and standing).

Flavia Benetazzo et al. [8] proposed, in 2014, a method to compute a respiratory rate that uses a RGB-D camera, to measure the breathing rate without any physical contact. In order to achieve the goals, the respiratory rate is computed by measuring morphological changes of the chest wall. This algorithm follows the following steps: detect the torso and shoulders, through the skeletal joints from Kinect, to find the chest of the user; define the ROI that is the chest (rectangle defined by the shoulders and a joint in the center of the torso); measure the depth changes in side of the rectangle of the chest; it is computed the mean value of the depth; calculate the weighted average of the four tests taken; and, finally, calculate the respiratory rate through the depth and the time of inhalation and exhalation. The system was tested with five healthy people, in stationary and moving modes. The results show that, in the worst case, the average mean of the error was 0.533 breaths per minute (bpm) and an average correlation coefficient of 0.9753. Thus, it is possible to conclude that the proposed algorithm can be used to measure the respiratory rate, both if the user is stationary or is moving. The principal limitation is that this system only works in frontal position (with an angular orientation not exceeding 25º).

Paul Hofland [15] proposed, in 2016, an algorithm based on the Kinect sensor that measure the breathing and heart rate without physical interfaces. The respiratory rate was obtained through the subject’s chest region movements that are captured by the depth camera from Kinect. The heart rate was obtained through the subject’s face, with the photoplethysmography method (that is the method used in an oximeter). The experimental work was performed on eight subjects and it was done fourteen tests with different operating conditions (sitting and standing positions) for each one. The results show that the respiratory rate could be determined although the algorithm, but sometimes fails to recognize some distances. The principal limitations of this work are the algorithm’s performance, that should be optimized, and the sampling rate was not consistent, so some data is missed. In this approach, the involuntary movements of the subject also are a limitation that are not solved.

Aleš Procházka et al. [16] proposed in 2016, a new method using Kinect sensor for monitoring breathing and heart rate without physical interfaces and to detect possible medical and neurological disorders. The data from Kinect was analyzed in two specify areas, the mouth and the chest area for measure the heart breathing rate. The proposed methodology assumes the use of digital filtering methods for data noise component rejection, resampling, data fusion and spectral analysis for detecting the required biomedical features. The results that were obtained verify the high correlation between the evaluation of the breathing frequency that was obtained from the image and infrared data of the mouth area and from the thorax movement that was recorded by the depth sensor. The achieved accuracy of breathing rate was 0.26% and the accuracy of heart rate estimation was 1.47% for the infrared sensor.
James M. Harte et al. [17] proposed, in 2016, a low-cost assessment system to monitoring the full chest wall motion and not just to measure the airflow of the lungs, as the spirometer do. This system aims to: identify motion at the chest; identify the effect of chest wall fluctuations; do the measurements to subjects unable to reliably perform a forced expiratory maneuver. This approach was developed using four Microsoft Kinect sensors (v1) to create a 3-D representation of a patient's torso over time. An evaluation of this system was performed in two phases. First to evaluate the accuracy of the system it was performed a static analysis. The evaluation object used to this first test was a training subject, of which volume could be computed. Initially, the subject was recorded by the Kinect-based system, five times during 15 s each, in a controlled room. After, a Nikon laser scanner was used to capture complete detailed geometric scan. With the results was estimated the volume of the body. Second, it was performed a dynamic (nine subjects with cystic fibrosis and 13 healthy volunteers). The acquisition protocol involved capturing analysis to measure the chest wall motion. It was used the developed Kinect-based system and the spirometer simultaneously in all participants quiet breathing for 20 s followed by a relaxed vital capacity maneuver and followed by twenty seconds of quiet breathing. The test was run three times per participant. From all acquisitions, it was obtained: tidal volume, total chest wall volume variation, respiratory rate and minute ventilation. The results of a static analysis showed that the system has slight underprediction of 0.441%. The results of a dynamic analysis showed an agreement with correlation coefficients above 0.8656 in all participants. The principal limitation is that the system has not the ability to measure the breathing rate on non-static subjects.

The aim of this work of measuring the breathing cycle without any intrusive device in the subject is a problem that had been studied at least since 2012. The Microsoft with a commercialization of a low-cost device (Microsoft Kinect) allows the researcher to develop new methods for numerous applications and the wireless measure of breathing rate is definitely one of them [8], [12]–[17], [20]. In these articles the approach taken is similar, they measure the variations of depth distances in the thorax area. These approaches showed very promising results, in general with a very low error between them results and the used as ground truth. The limitations of these approaches are the following: incapability of accurate measure of the breathing rate with a subject in movement ([12]–[17]), and the high level processing. The only approach that handles this problem uses the skeletal joints from Microsoft Kinect to track the chest of the subject [8].

In this thesis, the approach aims to present a solution to the high level of the processing issue, creating a solution that can run in a device with low processing cost.
2.3. Prism-based method

In this thesis work, it was used static points from Kinect skeleton. Therefore, in order to use this points in other ways, and not only use the raw signal from the axes to extract the breathing rate, it was searched for methods that could bring alternatives to our approach.

Carlos Massaroni et al. [21] proposed, in 2017, a novel method to compute breathing volumes changes and calculate respiratory parameters by using a motion capture system. This method is a prism-based method that computes the volume in the chest area by defining 82 prisms from 89 markers attached to the subject chest. The volumes are compared with the spirometer volumes and with the volumes computed by a conventional method based in tetrahedrons decomposition of the chest wall and integrated in a commercial motion capture system. The experimental was performed on eight healthy volunteers, it was collected thirty seconds of data in quiet breathing from each of them. The results show that the novel method have better values than the conventional method. The discrepancy of the novel method is 2.23 % (1.33 % lower than the one measured in conventional method) and a coefficient of determination equal 0.94 (0.02 higher than the one measured in conventional method). The sum of squared error of this method compared with the spirometer was 0.61 bpm, that was 0.07 bpm lower than the result achieved by the conventional method.

2.4. Difference of breathing between women and men

In an experimental work with non-chosen volunteers, there are a lot of parameters that can affect the study and influence the results. The gender of the participants is one parameter that can influence the results.

Maria Ragnarsdóttir and Ella Kristinsdóttir [22] proposed, in 2005, to establish reference data for breathing movements and patterns for healthy males and females that can affect the breathing movements. In this experimental work was used the Respiratory Movement Measuring Instrument (developed by ReMo), i.e. a convenient instrument measuring bilateral real time postero-anterior respiratory movements of the upper and lower thorax and the abdominal wall simultaneously. The experimental work was performed on 108 volunteers in a supine position. The subjects were instructed to breath normally and slowly during 60 s in each case. Symmetry, range, type, rhythm and frequency of breathing movements determined the breathing pattern. This study concluded that during deep breathing the women had significantly less abdominal movements than the men (less 6 %) and has been reported that males have about 20 % greater chest expansion than females, although the range of quiet breathing movements of men and women was not significantly different.
2.5. Breathing monitoring medical application

In various diseases it can be important to monitor the breathing cycle as a critical parameter. In the evolution of some diseases as cancer, particularly when it is not under medical control, the breathing rate can significantly change. Finding patterns in the breathing rate related with the status of a disease can be important to monitor the evolution of the disease. Adding the non-physical contact device as this thesis is searching for, it will improve the comfort of the patients, while being monitored.

Ahmad Abushakra and Miad Faezipour [23] proposed, in 2014, to develop a virtual environment to optimize breathing therapy for patients with lung cancer, by monitoring the subject’s breathing movements and motivating them to perform interactive regulative breathing exercises and at the same time provide a quantitative measurement of progress and compliance. This framework provides a 3-D computer animation of the human body where, for example, it is possible to analyze the evolution of the cancerous lung cells through certain breathing movements in this virtual environment. To capture the breathing movements, it was used the acoustic signal of respiration as one of the major inputs, in addition, the age, gender, height, and the cancer stage of the patient are also input factors of the framework. The framework is a smartphone application, that will integrate three components: breathing movement classification, lung capacity estimation and a visualization component. The patient is encouraged to regulate the breath with this application that shows the virtual lung organ/cell movements in real-time. In order to compute the lung capacity estimation through the acoustic signal of respiration, it was analyzed the energy of the signal to compute the forced vital capacity of the patient. To classify the breathing movements, it was split the signal into speech and silence segments, for each one it was analyzed some components of Mel-Frequency Cepstral Coefficients. In silence segments through a linear threshold, it identified if this segment is an inhale or an exhale. Lastly, this framework shows the animation of the lungs in real-time. The principal limitation is that this only introduces a conceptual framework, consequently, is not fully implemented.

2.6. Skeletal joints analyze

The approach presented in this thesis is based on the accuracy of the skeletal joints from the Kinect sensor to measure the breathing rate directly with these signals. In this section are reviewed two papers about the limitations of Microsoft Kinect sensor acquiring data and the most reliable skeletal joints to use. This research was made to provide a preliminary review of the skeletal joints accuracy to assist the present thesis.

Carlos Zerpa et al. [24] proposed, in 2015, a system to provide evidence of the reliability and validity of the Microsoft Kinect system measuring human movement analysis. In this study were tested twenty-six healthy participants (seven males and nineteen females). Two acquisition systems were used: the Vicon Peak Motus (VPM) human movement analysis system composed of two Basler FireWire cameras
and fifteen reflective passive markers, and a Microsoft Kinect camera to capture the 3-D human movement. The markers of VPM coincide with the landmarks that the Microsoft Kinect system automatically tracked. Both systems were connected to separate computers and synchronized, to start both systems at the same time. Each participant did two acquisitions, for five minutes, consecutively. The results show that the Peak Motus system have a high degree of reliability determined by the interclass correlation values. The joints of Kinect that had high reliability values are the following: left shoulder, right wrist, left knee, and right knee (as seen in Figure 2). The principal limitations are the small sample size of participants, small number of cameras for Peak Motus system and the tracking system of Kinect is unable to detect and track objects.

![Figure 2 - Reliability of displacement measures from the traditional and Microsoft Kinect system](image)

Mohamed Elgendi et al. [25] proposed, in 2014, to assess the arm speed movements via a Kinect sensor. During the experiment, it was recorded arm movements using a Kinect camera from 27 healthy volunteers. These collected movements were used as a benchmark for effective speed detection of an arm movement. All measurements were taken with subjects standing, initially with both arms extended along the sides of the body. Then the subject was asked to raise the dominant arm five times in three different speeds. The proposed arm movement classification types the algorithm consists of three steps. First, the pre-processing, where the coordinates and the instantaneous velocity of the points were computed and filtered by a low-pass filter. Then, the feature extraction, where it calculated the velocity and acceleration, and their mean and standard deviation. Lastly, it was done the classification with the most relevant features, the mean and standard deviation (SD) of the instantaneous velocity of the hand. With this, it was estimated threshold to classification. It was done a statistical Kruskal-Wallis and analysis of variance (ANOVA) tests, to investigate whether the hand-movement speed feature takes different values among the three different speed classes. In the case of interclass speed analysis, low p-values
were scored for both tests, which indicate a large difference in the means and medians of the three-speed classes. Analyzing the sample rate of the Kinect camera, it was discovered that this parameter is not stable, it fluctuates between 26.95 Hz and 33.67 Hz, instead of 30 Hz as it said in Kinect specifications. Mainly in a high-frequencies analysis, this can be an issue.
Chapter 3

Theoretical Background
In this chapter will be explained several concepts used in this thesis. To start, it will be explained the Breathing Cycle that is the base of this work, since the aim of this work is to detect the breathing rate through a novel approach. After it is introduced the Cross-correlation Algorithm and how it is possible to synchronize signals using it. This algorithm was used to synchronize the signals from Kinect and from Spirometer of the same acquisition. At section 3.3, it is explained how the depth sensor from Kinect works and how it is created the skeleton. Lastly, it is introduced the Spirometer, the ground truth device used in this work.

### 3.1. Breathing Cycle

In this work, the objective is to detect the respiratory movements through the detection of variations in the torso of a subject, finding a way to calculate the respiratory rate. It was done a research work to understand what the breathing cycle involves and where it is important to look for respiratory movements in the skeleton of the Microsoft Kinect One. The breathing function is briefly explained to fit this point at this thesis.

The Breathing Cycle is a description of the changes in pressure, lung volume and airflow. As shown in Figure 3, breathing is achieved by cyclically increasing and reducing the intrapulmonary pressure \( P_{ip} \) compared to the atmospheric pressure \( P_{atm} \), to the air flows from the area of higher pressure to lower pressure areas. The breathing cycle can be divided into three basic stages: rest, inspiration and expiration.[26]

![Breathing Cycle Diagram](image)

**Figure 3** Graphic Representation of pressure changes in lungs [26]

During inspiration, there is a net movement of air into the lungs and the volume of the lungs expands by the tidal volume above that of the Functional Residual Capacity, i.e. above the volume of air present in the lungs at the end of passive expiration. The diaphragm is a principal muscle that actuates in these
motions whose contraction pushes the abdominal contents downward, thus increasing the superior-inferior dimension, as well as pushes the abdominal contents outward, dragging the ribs outward and thus increasing the anterior-posterior dimension of the thorax.[26]

During expiration, there is a net movement of air out of the expanded lungs and the volume of the lungs declines by the tidal volume back to the Functional Residual Capacity. Expiration is achieved by contraction of the superior-inferior and anterior-posterior dimensions of the thorax. In the expiration stage the diaphragm and external intercostals relax, the highly elastic capacity of the lung lead to a natural recoil of them.

During rest, there is no net movement of air into or out of the lungs and the lung volume is equivalent to the Functional Residual Capacity.

As shown in Figure 4, these volume changes are responsible for breathing and are motivated by motion of the diaphragm and chest wall which combined aid to fluctuate the superior-inferior and anterior-posterior dimensions of the thorax.

Figure 4 Graphic representation of thorax extending during inspiration and thorax reducing during expiration

The number of respiratory cycles per minute is the respiratory rate and is one of the four primary vital signs of life. The normal range of a respiratory rate for a healthy adult at rest is between 12 bpm and 18 bpm, and can reach a range between 10 bpm and 30 bpm for a healthy elderly person at rest.

The respiratory rate is an important prognostic parameter to detect diseases, i.e. bronchial asthma, pulmonary embolism, and heart failure. That is because of hypercapnia (CO₂ retention), hypoxia (deprived of adequate oxygen supply), and metabolic acidosis (low pH in blood and tissues), lead to an increase in respiratory rate. Therefore, measuring the respiratory rate may support the early identification of high-risk patients. A published analysis of more than a million sets of vital data of patients in a United States hospital found an association between the deviation of the respiratory rate from normal
and hospital mortality of a size similar to that demonstrated for the external quality assurance pneumonia data. Thus, it is important to monitor the breathing rates to prevent major risks, and the abnormal respiratory rates should be seriously taken into account.[1]
3.2. Cross-correlation

In the experimental work were acquired data of Microsoft Kinect One and spirometer that was used as ground truth. It was not possible to acquire both signals simultaneously. Therefore, in order to solve this problem, it was used the cross-correlation between two signals to find the delay that occurred between the acquisitions and to be able to synchronize them shifting the spirometer's signal.

Cross-correlation is a measure of similarity between two different signals. The cross-function for discrete signals is defined for:

\[(f \ast g)[n] \equiv \sum_{m=-\infty}^{+\infty} f^*[m]g[m + n],\]

where \(f\) and \(g\) are the compared signals, \(f^*\) represents the complex conjugate of the signal \(f\) and \(n\) is the displacement (lag). The formula above computes the cross-correlation for a lag \(n\). To calculate the lag of two signals, it is important to compute the cross-correlation for all delays possible and store the values in an array. Once it is computed for all possible delays, it results in a cross-correlation series of the twice a length of the \(g\) signal plus the length of \(f\) signal. Founding the maximum value of cross-correlation in the returned array, it is found the lag between the signals. Consequently, it is possible to align the signals as it can be seen in the example below, Figure 5. In this work is used a MATLAB8 function called \(xcorr(x, y)\), that returns the cross-correlation of two discrete-time sequences for all lags possible.

![Cross-correlation example](image)

**Figure 5** Explanation of how synchronism works through cross-correlation. (Top) The signals misaligned; (Center) Result of cross-correlation between the signals above, in interval of lag [-4, 4] s; (Bottom) Signals align after correcting their delay. [27]
3.3. Microsoft KINECT (v2)

This research aims to test a reliability of the skeletal joints of Kinect v2 to measure a respiratory rate of a subject. The device to be used was the Microsoft Kinect v2, which is a human tracking peripheral used in the gaming industry for the Microsoft Xbox One that provide a low-cost 3D capture. The depth camera of the Kinect sensor returns a skeletal representation of the subject in front of it. The skeletal joints are frequently used to tracking people movement to use the body motion to play games in Microsoft Xbox and for this task is not crucial the accurate of joints positions. [28] Those signals are noisy and the error due to the way how those signals are created.

3.3.1. Time of Flight

The Kinect v2 sensor is based on Time-of-Flight (ToF) principle, that refers to the process of measuring depth distances of the scene. This measurement is made through quantification of the changes of an emitted light signal hitting an object in a scene and returning. [29], [30]

The principle of ToF imaging is measuring distances to a 3-D object by measuring the absolute time that a light pulse needs to travel from the source (sensor) into a 3-D scene and return, after reflection. The speed of light is constant and known, \( c \approx 3 \cdot 10^8 \text{ m/s} \). As can be seen in the Figure 6, this method count the time between emission and detection the light pulse, to measure distances. [29], [30]

![Figure 6 Schematic of how the ToF method works.][31]

The sensor estimates in real-time a distance value to all scene and these distances are directly stored in a matrix of 512 x 424, the result of this is called a depth map. [29]

E. Lachat et al. [29] testing the influence of the brightness conditions on the measurements. This experiment shows that the sensor can work during a sunny day only if that the light does not directly illuminate the sensor. However, during a sunny day acquisition, two phenomena can happen: the number of “flying pixels” increases particularly on the edges of the sensor field of view and the number of points decreases with the light intensity.
3.3.2. Skeleton Tracking

Skeletal tracking algorithm allows Kinect to recognize people (maximum of six persons) and track their movements (maximum of two persons simultaneously). The face recognition is only possible in some face positions, as it is possible to see in Figure 7.

This algorithm is optimized to recognize users standing or sitting and facing the Kinect.

![Figure 7 User face position for it to be detected](image)

Using the infrared camera, it is possible to calculate the person's depth in the field of view allowing the recognition of different persons and different body parts.

For best tracking results, it is important for the target to be in the default range. The default range for depth is between 1.2 m and 3.2 m, for height 43.5 degrees vertical Field of View (FOV) and for width 58 degrees horizontal FOV. In the Figure 8 it can be seen the default range (depth, height and width).[32]

![Figure 8 Default range of camera](image)
3.3.3. Skeletal Joints

Jamie Shotton et al. [28] proposed, in 2011, the method used to predict 3-D positions of body joints from a single depth image quickly and accurately to be used in Microsoft Kinect. As illustrated in Figure 9, the Kinect create the skeletal representation in a three-stages process. First it is computed a depth map using a ToF algorithm. Then, to infer body position, it is used a machine learning algorithm, that is randomized decision forest, mapping depth images to distinguish body parts. To transform the body part image into a skeleton, it is used the mean shift algorithm to robustly compute modes of probability distributions, that generate confidence-scored 3-D proposals of several body joints by re-projecting the inferred parts into world space. To do that, it is localized spatial modes of each part distribution and thus generate confidence-weighted proposals for the 3-D locations of each skeletal joint. This system runs at 200 frames per second on XBOX Graphics Processing Unit.

![Image](image_url)

**Figure 9** Steps taken to compute the skeletal joints. From an input depth image, is inferred the body parts (Colors indicate the most likely part labels at each pixel, and correspond in the joint proposals), and the 3D location of the body joints are estimated.[28]

The density estimator used to estimate the body part is defined as

\[ f_c(\hat{x}) \propto \sum_{i=1}^{N} \omega_{ic} \exp \left( -\frac{\| \hat{x} - \hat{x}_i \|^2}{b_c} \right) \]

where \( \hat{x} \) is a coordinate in 3-D world space, \( N \) is the number of image pixels, \( \omega_{ic} \) is a pixel weighting, \( \hat{x}_i \) is the reprojection of image pixel \( x_i \) into world space given depth \( d_i(x_i) \), and \( b_c \) is a learned per-part...
bandwidth. The pixel weighting $\omega_{ic}$ considers both the inferred body part probability at the pixel and the world surface area of the pixel:

$$\omega_{ic} = P(c|I,x_i) \cdot d_i(x_i)^2$$

The joint prediction accuracy can be seen in Figure 10. The average precision results on the synthetic test set, achieving a mean average precision of 0.731. In the Figure 10 it is compared a performance of the system (red) with the best achievable result (blue) given by ground truth body parts.

Figure 10  Average precision of 16 joints from Kinect skeleton[28]
3.4. Spirometer

The Spirometer is a precision differential pressure transducer for measurements of respiration flow rates. Basically, the Spirometer and attached flow head work together as a pneumotachometer, with an output signal proportional to airflow. This instrument is a device contained in the Exercise Physiology System of ADInstruments (that include LabChart Pro 8 software, Gas Mixing Chamber, Spirometer, Thermistor Pod and Exercise Physiology Accessory Kit that have the face mask and other accessories as can be seen in the Figure 11), based in a complete physiology recording system for monitoring cardiorespiratory and metabolic function. The system records and shows continuous real-time measurements of metabolic parameters, such as CO2 and O2 concentrations, airflow, temperature of respired air, ECG or EMG.

Figure 11 – Materials contained in the Exercise Physiology System of ADInstruments
Chapter 4

Methodology
In this chapter it will be explained the various stages of the experimental work carried in this thesis. In Figure 12, it can be seen how the project was divided. The seven presented modules will be individually explained in the respective section.

Figure 12 - Block diagram of the methodology
4.1. Experimental Protocol

In order to solve the purpose objectives, the first part was to acquire data from Kinect and from Spirometer in 28 volunteers. Briefly, the data extraction procedure can be seen in Figure 13.

The initial step was to recruit volunteers, to make the acquisitions. The volunteers were recruited in the Instituto Superior Técnico Campus and, they were not all students, researchers were also recruited. The effort to recruit older volunteers was made to increase the spectrum of ages. All participants were advised to bring sportswear to prevent participants from transpiring their daily clothing. The experimental data was collected from 28 healthy subjects (23 males, 5 females; 24.86±3.15 years old, height 176.46±8.08 cm; weight of 73.18±10.24 kg). The participants responded to informed consent where they all understood the meaning of the study and authorized the collection, recording and processing of data. (Appendix A  Informed Consent)

To better analyze the performance of the Kinect points on measuring the breathing rate, it was decided to make sixteen acquisitions for each participant, eight in rest mode and eight after-exercise.
modes. In addition, for each mode it was made half acquisitions seated and the other half standing. All sixteen acquisitions were performed for all the participants and lasted one-minute each one.

In order to avoid failures, a guide was followed for this study. First, it was explained the study, the objectives, and the procedures. Then the participants read and responded to the informed consent. After that, to begin it was carefully placed the face mask in the subject in order to prevent air leaks. At that point, it was made the first acquisitions, four one-minute seated acquisition followed by four one-minute standing acquisition. The next step was to make the acquisitions after exercise. The exercise was divided into 3 stages: to begin the participant walked one minute at 6 km/h, the next minute the speed was increased to 8 km/h and a slope of 10º, finally the speed and the slope was increased adapting them to the current participant and was maintained for one minute. Depending on how the subject recovers from exercise, it was made one or two acquisitions between physical exercises. This evaluation was based on the level of tiredness, i.e., if the participant was still breathless, the second acquisition was made without repeating the physical exercise. During all acquisitions, the participants used a face mask from Exercise Physiology Kit of ADIntrument\textsuperscript{TC}, as can be seen in the Figure 14.

![Figure 14 - Face mask used in the experimental study.](image)

All participants were positioned at the same distance from Kinect sensor. The Kinect was placed on a specific carrier, it means 95 cm from the floor. On the seated tests, the chair was aligned with the z axis of the Kinect sensor and it was placed 215 cm from the sensor. On the standing tests, as well as the seated tests, the participant was aligned with Kinect and it was placed 215 cm from the sensor. In Figure 16, it can be seen two photos of the acquisition space, with one participant during seated and standing acquisitions. In Figure 15, it can be seen a photo of a participant running on the treadmill of the laboratory.
Figure 15 - Photo of a participant running on the treadmill.

Figure 16 – Photos of how the acquisitions were made in the laboratory. (left) Seated acquisition; (right) standing acquisition.
4.2. Experimental Setup

In order to test the reliability of Kinect skeleton points on measuring the respiratory rate, it was decided to compare the data from the Kinect skeleton with the data from Spirometer. The data from the spirometer was acquired from the laboratory’s computer through the LabChart, a software from ADInstruments. The data from Microsoft Kinect v2 was acquired and saved with a MATLAB code developed in MATLAB R2016a and run on a PC with an Intel(R) Core(TM) i7 CPU 4710HQ @ 2.50 GHz processor with 12 GB of RAM. The data acquisition was performed at the same time but not synchronously.

The experimental work was done at the Lisbon Biomechanics Laboratory which is located at the Instituto Superior Técnico. This laboratory has the ideal conditions to support this study, in addition to the spirometer equipment it has one treadmill (Woldmedica Marathon Medical that has medical certificate), where the participants could perform the physical exercise before the acquisitions in which they supposed to be under physical effort, in order to increase the breath rate. Furthermore, it was possible to control the brightness of the room, as well as its temperature.

4.2.1. Microsoft Kinect v2

The Microsoft Kinect v2 was used as a depth camera which computes a skeletal representation of the human body, which consists of 26 notable points (joints and body extremities), as it can be seen in the Figure 17. Since the objective of this work is to measure the respiratory rate, it had to be chosen the most significant points for this purpose.

Figure 17 - Skeleton positions relative to the human body [29]
It was concluded in a preliminary work that the areas of the body that are not possible to extract any breathing movement are the arms and the legs. When a subject breath, the legs are not affected by this action, and the arms are much more sensitive to involuntary movements than the shoulders. In addition, the head points were also excluded for the same reason as the arms. Thus, the most significant points are: both shoulders and three points at the center of the torso. The five skeletal points used are the following (as seen in Figure 18-left): 1) spine base point, 2) spine mid point, 3) left shoulder point, 4) right shoulder point and 5) spine shoulder point.

The movements of the points were observed in 3 dimensions. In the x-axis it can be seen the lateral movement, in the y-axis it can be seen the vertical movement, and in the z axis can be seen the depth movement (towards the Kinect). In the right side of Figure 18, there is a graphic representation of a referential used in each axis of each skeletal joint, where it can be seen the direction of each axis, using the direction of the sensor as reference.

Theoretically, the best points will be located higher up. However, for a more abdominal breathing, the lower ones can produce better results [20], [22]. Therefore, the points 3, 4 and 5 should better detect the breathing movement, although for a more abdominal breathing the points 1 and 2 should detect it in a better way.

![Figure 18](image)

**Figure 18** – (Left) Depth map from Kinect with the five joints used marked in the image in red; (Right) 3-D axis orientation representation with the location of the Kinect sensor as reference.[33]

### 4.2.2. Spirometer

In this experimental work, the feature used from the Spirometer data was the airflow signal. Through the airflow data, it was possible to compare this data with the fluctuations of the skeletal joints from Kinect. As it can be seen in the Figure 19, the face mask was coupled to a chamber with a valve system that allows breathing in from outside and exhaling to the interior of a mixing chamber (Gas Mixing Chamber) where the different sensors (temperature, O$_2$, CO$_2$, and module) are connected. Finally, this chamber is connected to the spirometry module which, in turn, is connected to the laboratory computer.
The sampling rate in the spirometer acquisition is 1000 Hz but, so that the signals would be easily processed/compared and to reduce the data size, the sample is converted to 30 Hz. The resample used to convert the sample rate applies an anti-aliasing lowpass filter (Finite Impulse Response) to the spirometer signal and in the end, compensate the delay introduced by the filter.

Figure 19 – Graphic representation of how the Spirometer device was attached to the remaining devices.[34]

In order to acquire the signal from spirometer, the first step was the calibration. The calibration of the spirometer was performed by injecting three liters of air with a calibrated 3-liter syringe (from Hans Rudolph), at low flow (1 to 2 liters/second), medium flow (4 to 6 liters/second) and high flow (8 to 10 liters/second). This is a very important procedure that can never be forget before the first acquisition of each participant test the system.
4.3. Pass-band filter

The starting point for the analysis of the breathing rate was using the EVM algorithm [2] to detect the breathing movements (Appendix C). This approach was rejected because this method filters the signal temporally and spatially and the spatial filter is unnecessary in the context of the current thesis problem. However, it was used the ideal pass-band filter developed and used in the EVM algorithm. Filtering the skeletal joints signals was essential to denoise and isolate the breathing movements in these signals.

It was used fifteen signals from skeletal joints of Kinect, the three axes of each of the five selected joints. These fifteen signals were filtered with the passband filter with a lower cutoff frequency in 5/60 Hz (5 bpm) and a higher cutoff frequency in 40/60 Hz (40 bpm). In Figure 20, it can be seen an example of a signal in raw (left) and a signal after the filter is applied (right).

A typical respiratory rate of a healthy adult at rest is between 12 bpm and 18 bpm. However, the group of participants included athletes and the participants were tested after exercise, so the bandwidth must be higher. Some preliminary tests helped the choice of the cutoff values and it was possible to verify that the typical respiratory rate values are not a good option as cutoff frequency. It was seen that for some rest tests and for most of the post-exercise tests, the filter cuts the frequency of the participants breath using a bandwidth between 12 bpm and 18 bpm. For example, it was found that for some subjects the respiratory rate was around 6 bpm resting and others can increase the breath rate to values above 30 bpm after exercise.

![Unfiltered signal (Point 5y)](image)

![Filtered Signal (Point 5y)](image)

Figure 20 - Example of a signal before and after the filtration. (Left) Unfiltered signal, (Right) Filtered signal
4.4. Methods

For the experimental setup data, it was chosen alternatives to the raw points’ axes method (first method). The first method analyzes the variations of the axes of the skeletal. The second one analyzes the variations of barycenter's axes of the 7 triangles (illustrated in Figure 21) created with the points of the first method. Lastly, the third method analyzes the variations of the area of the same triangles used in method two, during the acquisitions. The second and the third methods were chosen to obtain more stable signals, with less noise and less sensible to involuntary movements of the participant. These two methods are inspired on the method proposed by Carlos Massaroni [21] that measure breathing parameters as breathing volume and breathing rate. That method uses 89 markers attached in the subject to create prisms. The second and third method in this work used a similar approach, it uses the five skeletal joints as the markers that Carlos Massaroni used in his work. The created triangles and his respective identification can all be seen in the Figure 21.

Figure 21 - Graphic representation of the triangles used in the methods two and three

The second method analyzes the variations of the central point of the triangles explained above. The purpose of that was to reduce the three points in only one point (the barycenter). It was expected that this center point was sensible to the same variations of the three vertices and less sensible to the noise of the isolated point, it means, to be able to better detect the breathing rate.

On the other hand, the purpose of the third method was to analyze the variations of area of the triangles and it was expected that the involuntary movements of the participant during the tests was eliminated, once it does not analyze the axes of points that are more sensible to those movements. Although the request to hold position for 1 minute, the involuntary movements were recurrent during the acquisitions because it was not easy for the participants to hold their position, especially after vigorous exercise and with a face mask that hinders the normal breathing. Using the variations in the area's size of the triangles to measure the breathing rate of a subject, it should solve the measurement problem that arises when the subject moves since the variations of the area must not be affected by those
movements. In contrast, the other two methods that extract the breathing rate from the spatial variances along the three axes in the skeletal points (method 1) and the barycenter of the triangles (method 2), the movement of the subject’s problem affects them.

4.5. Cross-correlation

The cross-correlation stage was done to synchronize the signals. As said before, the spirometer data and the Kinect data were not acquired simultaneously. In order to accurately compare the results from these two different systems, it was imperative that the signals were synchronized. The first step to do at this stage was to compute the cross-correlation along the time, between -6 s and 6s, since the delay/advance between the beginning of the acquisition was never higher than 6 s. The lag of the systems is reached finding the max value of the cross-correlation array. Once calculated the lag value, it was possible to synchronize the Kinect signals with spirometer signals.

Further in this section, it is presented one example of the synchronization performed for each one of the three methods used. It can be seen how the signals align after synchronization procedure and how similar are the signals from the three methods and from the spirometer.

The Figure 22 is an example of a synchronization between a signal from the first method and a spirometer, before and after the alignment. The maximum cross-correlation value was, in this case, 336,78 and the lag between the signals was -0.60 s.

Figure 22 – Alignment between a y axis of spine shoulder point and the spirometer signal. (Left) Unsynchronized signals; (right) Synchronized signals.

The Figure 23 is an example of a synchronization between a signal from the first method and a spirometer, before and after the alignment. The maximum cross-correlation value was, in this case, 143,17 and the lag between the signals was -3.33 s.
The Figure 24, is an example of a synchronization between a signal from the first method and a spirometer, before and after the alignment. The maximum cross-correlation value was, in this case, 115.52 and the lag between the signals was -2.07 s.

Figure 23 - Alignment between a x axis signal from triangle 1 and the spirometer signal. (Left) Unsynchronized signals; (right) Synchronized signals.

Figure 24 - Alignment between a signal of the variations of area from triangle 3 and the spirometer signal. (Left) Unsynchronized signals; (right) Synchronized signals.
4.6. Calculation of Breathing Rate

After the synchronization, the signals were split in half to increase the number of data samples. This way, it also allows checking if the system works with samples of just 30 s and, consequently, if it measures accurately in 30 s. The measuring process in 60 s should be more accurate since the size of information is twice larger. Then, with the signals synchronous and divided it was analyzed the Fast Fourier transform (FFT) spectrum for all signals (the half-signals).

During the acquisitions, the breathing rate of a person can change along the time. Therefore, the frequency of the maximum value in the FFT spectrum may not be the average frequency of that signal. This problem remains even with the 30 s signals, that are the half-acquisition time, because the breathing rate could change at any second. In order to solve this issue, it was used the 3 higher peaks from the FFT, and it was computed the Weighted Arithmetic Mean (WAM) to calculate a mean frequency of these three peaks, using the peak values as weights. It was only used three peaks because, increasing the number of peaks used, it also increases the risk of include peaks that are not related to the breathing movement causing a higher error between measurements. The WAM can be calculated using the following equation:

\[
\bar{x} = \frac{\sum_{i=1}^{n} \omega_i x_i}{\sum_{i=1}^{n} \omega_i}
\]

where the 3 peak values are the \( \omega \) of the WAM equation, the \( x \) is the frequency of each maximum and \( n \) the number of peaks that are three. Then, with the frequencies computed, to finish, it was calculated the error between the method’s signals and the spirometer signals.

4.7. Evaluation Metrics

To evaluate the system, it was compared the breathing rate obtained from the three Kinect methods explained in section 4.4 and from the ground truth (the spirometer data), subtracting the values to calculate the error. In the data set, there are results from clean acquisitions and for acquisitions who have experienced problems. These problems range from the oscillated subject during the acquisition to air leaks from the face mask.

In order to exclude the outliers, it was followed several steps. First, for each one of the four modes, it was computed which method obtains the best results. For that purpose, it was computed the median of the error for each method. The method that obtained the lowest median of error was used to reach the outliers. That is because the method that obtained the lowest median is expected to have the most reliable results. Consecutively, it was computed the median of the errors in all signals from this method, for each acquisition. The acquisition data that had a median error higher than 5 bpm was considered outlier (as seen in Figure 25). The threshold value (5 bpm) was high to only consider as outlier the worst
results. It seems important that the final data has not perfect results to be possible analyze the system in a realistic way.

![Histograms of median values between axes of method 1 along all acquisitions.](image)

**Figure 25 - Histogram of the median values between the axes of method 1 along all acquisitions.**

While exist axes that present a global better result, the best way to measure the breathing rate was to detect the best axis for each acquisition. Although it is possible to find an axis that presents better global accuracy for each mode, the acquisitions are independent and the best axis for each acquisition can be different. So, with an automatic detection of the best axis in each acquisition, it will certainly improve the accuracy of the system, when compared to the one that uses the best axis for each mode.

To achieve the best axis, for each acquisition, it was tested four methods: 1) the maximum amplitude of the filtered signal, 2) the maximum peak value of the FFT of the filtered signal, 3) the maximum sum of the three higher peaks of the filtered signal and 4) the maximum mean of the three higher peaks of the filtered signal. In all methods, the axis that had the higher value in each acquisition was used to compare with the spirometer signal (ground truth).
Chapter 5

Results
As explained in the section 4.7, the outliers were deleted from the data. Thus, from 896 data samples only 520 data samples remain, 173 of them (77.23 %) from the rest seated tests, 155 (69.20 %) from the rest standing tests, 119 (53.13 %) from the seated tests after exercise and 73 (32.59 %) from the standing tests after exercise. Therefore, the final data sample contains 58.04 % of the initial data.

To choose the most reliable signal to measure breathing rate, it was processed all signals from Kinect with the Spirometer signals (ground truth) and it was computed the error between them. In figures 26, 27 and 28 it is possible to see the mean and the standard deviation of each signal in the four different experimental conditions. Through this graphics, it is possible to observe the signal with less error, i.e. the most reliable signal, for each mode of acquisition.

In Figure 26, it is notable that in the seated acquisitions (left graphs) the z-axes presented lower error, and the x-axes higher. In the standing acquisitions (right graphs) the axes that have the lower error is the y-axes, although the higher error change for each joint, in some cases it is the x-axes, in others it is the z-axes. Every difference is significantly visible in the graphs of the acquisitions after exercise (bottom graphs).

Figure 26 - Mean and Standard Deviation in the method of points axes in the four modes of acquisition. (Review joint corresponding to the number of the horizontal axis in Figure 18-left)
Figure 27 - Mean and Standard Deviation in the method of Barycenter axes in the four modes of acquisition. (Review triangle corresponding to the number of the horizontal axis in Figure 21)

Figure 28 - Mean and Standard Deviation in the method of the triangle area in the four modes of acquisition. (Review triangle corresponding to the number of the horizontal axis in Figure 21).
In Figure 29, it can be seen in four images the heat map of the joints for each acquisition mode. The circles are a representation of the axis, where each one of the five chosen joints has three independent axes, that are grouped in the images. The axes in top correspond to the y-axes, the axes in the right correspond to the x-axes and the other corresponding to the z-axes. The direction of each axis can be seen in Figure 18-right, the circle's position has the same orientation as that.

![Figure 29 - Heat map of the error along all tested axes.](image)

In Figure 30, it can be seen the graphics of the frequencies obtained from the spirometer and from the best axis for each mode along the acquisitions without the outliers. The choice of the best axis for each mode was made through the axis that obtained a smaller mean error. In those graphs, it is possible to visualize the difference between the ground truth and a chosen axis for each acquisition mode, along the inliers used.
In order to analyze the reliability of our approach to measure the breathing rate, without removing any outliers, it was created the graphic presented in Figure 31. The graphic shows the mean and the standard deviation of the error for each acquisition mode (blue) and, it can be seen the median also for all four modes (red). In Table 1, it can be seen the values of the graphic present in Figure 31.

The methods tested in Evaluated Metrics stage to find a way to automatically choose the best axis for each acquisition does not had positive results. Although it was not possible to find a method that automatically choose the best axis to use to measure the breathing rate, in Figure 31, it can be seen how good the results of this method are with all acquisition, also including the bad ones. Analyzing the median, it is possible to see a draft of the results without “outliers”. It is important to analyze it without outliers because during an acquisition many external factors can influence the results, since the clothes thickness till the involuntary movements of a subject.

Figure 30 - Graphics of the breathing rates expected from Kinect joints and from Spirometer.
Table 1 - Results of the best axis for each acquisition for each mode.

<table>
<thead>
<tr>
<th></th>
<th>Rest Seated</th>
<th>Rest Standing</th>
<th>Seated After Exercise</th>
<th>Standing After Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MEAN (BPM)</strong></td>
<td>1.07</td>
<td>1.43</td>
<td>1.26</td>
<td>3.04</td>
</tr>
<tr>
<td><strong>SD (BPM)</strong></td>
<td>1.92</td>
<td>2.21</td>
<td>1.71</td>
<td>3.62</td>
</tr>
<tr>
<td><strong>MEDIAN (BPM)</strong></td>
<td>0.35</td>
<td>0.48</td>
<td>0.62</td>
<td>1.43</td>
</tr>
</tbody>
</table>

If the axis used to measure the breathing rate could be automatically selected as the best for each acquisition (without removing the outliers), the percentage of error would be different to the percentage of error of our first method (without outliers). It can be seen this comparison of the percentage of error in Figure 32.
Figure 32 – Comparison of the percentage of error between method 1 without outliers and the results using the best axis for each acquisition.
Chapter 6

Discussion
In this section, it is discussed the results obtained in this thesis and compare them with the similar approaches mentioned in section 2.2. Starting with the analysis and comparison of the methods used to determine respiratory frequency, it can be seen in Figure 26, Figure 27 and Figure 28 that the method with better results is the first one, the one that uses the raw skeletal joints axes. The mean error of the axes from the skeleton joints (first method) are around 2 bpm, in contrast with the results from the second and third method that are around 7 bpm.

In the results from seated acquisitions with the first method, the x-axis of the points clearly has the greatest errors. This was expected given that during the breathing cycle the thorax volume changes in the superior-inferior (y-axis) and anterior-posterior dimensions (z-axis), as it was described in the section 3.1. the z-axes had the best results when compared with the results of the other axes. The participants of this experimental work were mostly from males (82 %) and this fact can explain the smaller error of these axes. As Maria Ragnarsdóttir and Ella Kristinsdóttir [22] conclude in its study, the males use more the abdominal area (z-axis) in deep breathing, while females use more the thorax area (y-axis).

The results from standing acquisition with the first method show that the smaller errors were noticed from y-axes and the higher errors from z-axes or from x-axes depending of the joints. Although it was not expected, in some joints, the z-axes were worse than the x-axes. The x-axes correspond to the side movement, and during a breathing cycle the movement should be manifested along the y and z-axes, as explained in section 3.1. (using as reference the directions described in Figure 18-right).

The biggest noticed difference between rest and after exercise acquisitions (visible in the results of first and second method) it was that with the increase of breathing rate and, consequently, the amplitude of the respiratory movements, the discrepancy between the accuracy of the axes are most visible.

The second and third methods tested did not have any better results than the results obtained in the first one. In the second method (Barycenter), it can be concluded that the best axes for seated acquisition are the z-axes and for standing acquisitions the y-axes are the axes with less error. However, this method was presented worse results than the first method, with the mean errors around 6 bpm, in the better cases. In the third method (triangles areas), the results were not any better than the results from second method.

Observing all the results from the three methods, it can be concluded that the first method obtained better results. In this method, for the seated modes, the best axis was the z-axis from the left shoulder. In the standing modes, the best axis was the y-axis also from the left shoulder. The fact that the left shoulder had better results coincide with the study done by Carlos Zerpa et al. [21] in 2015, where it is concluded that the left shoulder is one of the most accurate joints in the Kinect skeleton. Although, these results contradict the accuracy results from Jamie Shotton and Toby Sharp [28] in 2013, where the right shoulder had slightly better results than the left shoulder (as seen in Figure 10).

It was not achieved a way to automatically select an axis (from Kinect data) that had better results measuring the breathing rate, for each acquisition. Although, observing the Figure 31, it can be concluded that, if it could be arranged a way to automatically choose the best axis for each acquisition, the results would be significantly better. The results using only the axis that have results closer to the
results from spirometer show that would be a great asset to achieve a method that finds the best axis for each situation. In the graphic of the Figure 31, it can be seen the median of error between this axis and the result from spirometer is around 0.5 bpm in the three first modes (rest seated, rest standing and seated acquisitions after exercise). In the standing after exercise acquisition, the median of error was not close to 0.5 bpm but it was under 2 bpm. Analyzing the median, it can be predicted how the results would be without the outliers (bad acquisitions). Since the median values are significantly lower than expected, it may be concluded that the percentage of error could be smaller if it was used the best axis for each acquisition, automatically selected.

Even removing the outliers (method 1), it was noticed that the error of the better axes was around 2 bpm. The error present could be due to the instability of joints from Kinect [28] and also due to the sampling rate of Kinect [25]. Microsoft mention that the Kinect sensor has a sampling rate of 30 Hz, but the sampling rate can fluctuate between 29 and 33 Hz (according to Mohamed Elgendi 2014 [25]). If we are considering that the sampling rate is 30 Hz and it is being 33 Hz, it can change the signal frequency from 12 bpm to 13.2 bpm and from 18 bpm to 19.8 bpm.

In addition to the technical potential problems referred above, during the acquisitions there are various factors that can affect the results. One of the principal factors are the thickness of the clothes, that can influence the movements in the depth image and consequently the movement of the skeletal joints. Another factor that certainly influences the results is the involuntary movements of the participants (staying steady for 1 minute with a spirometer face mask was not an easy task). It was only tested one method that could deal with it, the third method, which unfortunately did not present good results. The mean of the error between the frequency measured with the spirometer and the results from this method was around 6 bpm. The large error obtained with this method precludes the use of this in future work.

During the acquisitions, the breathing tube that connects the face mask with the gas mixer chamber naturally stays positioned in front of the chest. In order to minimize the influence of the tube in the selected joints movements, the tube was placed beside the central torso joints. However, this can also influence the joints positions, since the joints position is computed analyzing the depth map of the area the tube causes noise, as explained in section 3.3.3.

In comparison with similar prototypes, our three methods present worst results, also comparing with our first method, that was the one that had better ones. In this case (first method), the mean of error was around 2 bpm. Although, if we could only use the results of the best axis for each acquisition, automatically selected, the mean of error would be significantly inferior (as seen in Figure 31). So, it can be concluded that this novel approach has potential, but is essential to discover a method that automatically choose the appropriate axis to be used to measure the breathing rate for each independent acquisition to compete with the similar approaches.

The similar prototypes [8], [12]–[17] use essentially the depth image directly, isolating the ROI that represents predominantly the torso area, and measure its fluctuations to compute the breathing rate or the lungs volume variation [12], [13], [15]–[17]. However, there are other contactless approaches like Natascia Bernacchia et al. [14] in 2014 that only use the abdominal area, Ahmad Abushakra and Miad
Faezipour [23] in 2014 uses the acoustic breathing signal captured by a smartphone to predict the evolution of the cancerous lung cells. These approaches showed accurate results, in general, with a very low error between the results and the used ground truth. The proposed system showed promising results with the use of the skeletal joints. The percentage of error of the results using the best axis for each mode was significantly higher than the related work. Although, analyzing the results of the best axis of each acquisition, the accuracy was similar to the related work that shows the importance of creating a method that automatically chooses the best signal to use in the breathing rate measurement.

It was possible to identify two limitations in the method proposed in this thesis that are also present in almost all similar prototypes. The system only works with static subjects and the thickness or loose clothes can interfere the measurement performance. An exception is Flavia Benetazzo et al. [8] approach that uses the skeletal joints to define the ROI, in order to handle the movement of the subjects. The limitation of the thickness or loose clothes is a common limitation in the prototypes that use depth cameras and measure the breathing rate computing the chest fluctuations. However, it can be asserted that with minimal data processing it was achieved very promising results.
Chapter 7

Conclusions and Future Work
The breathing rate is an important parameter to track in order to prevent serious complications in patients. Measure this parameter without any physical interfaces can significantly improve the health screening for example in waiting rooms [7].

The results that were obtained with the presented approach show that it is possible to detect breathing movements from the skeletal joints acquired from Microsoft Kinect v2. It was analyzed the accuracy of each axis of the skeletal joints in measuring the breathing rate, and it was possible to conclude that, for seated acquisitions, the z-axis of the left shoulder obtained a mean error lower than the other axes and, for standing acquisitions, the y-axis of the left shoulder obtained a mean error lower than the rest.

The percentage of error of this system, using the axis that presented an average of error lower, was higher than the similar approaches reviewed. However, if compared with the best signal for each acquisition, the percentage of error was similar. This factor deserves some future work due to his promising results.

This approach differentiates from related work since it uses signals that are already processed automatically by Kinect, in contrast with others that computes the variations of chest distances to the sensor. This approach has achieved interesting results but still needs improvements. This system gains an advantage by achieving results with a minimal data processing, which can be useful to apply in low processing devices or to run in the background without causing any negative processing impact.

This system can be used in waiting rooms of a hospital or clinic to detect variations of the breathing cycle of a patient and alert the nurses or it can also be used in the gaming industry. Since the device used is a Microsoft Kinect, that is available for XBOX gamers, this system can act in order to make the games more interactive, using the breathing rate as an input, so that the game can adapt, becoming more stimulating for the player. It can also be used in augmented reality, using this system to measure the breathing rate of the subjects in a room and display it in an augmented reality device [35].

At the beginning of this work, it was set three main goals. The first two was determine if the skeletal joints from Kinect data have any physiological information and to conclude if it is possible to measure the breathing cycle from the skeletal joints. Although the accuracy of the first method was not better than the similar approaches, it has been proved that it is possible to measure the breathing rate through the skeletal joints. The third goal was to analyze the reliability of these signals, validated with a more accurate device. This third goal was also accomplished. It was found the most accurate skeletal signal for each acquisition mode and his respective value.

**Future Work**

The explored algorithms and methods leave room for improvement and open up interesting investigation avenues, which are proposed in this section. This work shows that the skeletal joints from Microsoft Kinect can be used to measure the breathing rate. However, to turn this into a commercial product it is essential that this project continue. For future work is important to address the follow goals:
• **Test the system in transitions of breathing rate** – The changes of the breathing rate are an important parameter to study; these changes can prevent serious problems in hospitals patients with an early notification to nurses or doctors. In the gaming industry, these changes can use the excitement status of the gamer as an input; that feature could allow the games to adjust to the individual.

• **Develop a method to choose the best axis for each acquisition automatically** – In order to compete with the similar approaches, it is important to develop a method that chooses, automatically, the best axis for each acquisition. In this work, it was concluded that the accuracy of the system could be improved if it was used the best axis. A machine-learning algorithm could be used to solve this problem.

• **Develop a method to work with moving subjects** – In a real situation, people move and oscillates. So, to measure the breathing rate without physical interfaces attached is essential the creation of a method to our system that can distinguish between breathing movements and other movements. An option to solve this problem, without commit the performance, could be use all skeletal joints to detect the subject’s movement.

• **Real-time application** – This type of systems aims to run in real-time to be useful. The last part of this project should be the development of a real-time application for medical purposes or for the gaming industry.

This research shows one of the applications that the Kinect v2 is capable of. The SDK 2.0 from Microsoft allows programmers to investigate the possibilities of the Kinect v2 sensor, which greatly facilitates the research in this area.
References


contact measurement of heart and respiration rates based on Kinect™,” pp. 4–8, 2014.


 Appendix A Informed Consent

No video or audio recordings of subjects were made in this study. The Kinect sensor provided numerical data that directly related to the position of each 3D point of the torso. Only this numerical data was stored in the database. The volunteers completed an informed consent, in which everyone agrees with the study and provided the following data, height, weight, age, and gender. It was confirmed that all volunteers were healthy, all responded that they did not have any respiratory disease. Below it is presented the informed consent used and the respective questionnaire that was answered by all the volunteers who participated in this work.

Dear participant,

We are conducting a study on the ability of 3D points of the Kinect v2 skeleton to reliably calculate the resting rate of an individual at rest and after physical activity.

To do this, we need your contribution!

The activity should take almost sixty minutes, during which you can perform several important experiences.

The acquisitions will be made through a Microsoft Kinect v2 sensor, which will only be used the camera's depth data. All data collected will be kept confidential and it will be analyzed exclusively by the researchers of this project. The data may also be used to present or display results in scientific publications, conferences, or similar events. The collected data can be used for online dissemination (for example, on YouTube, Vimeo, etc.) only after your specific authorization.

Your participation is voluntary and you can always give up at any time without any penalty or consequence.

To participate in this experiment, we ask you to fill out the consent form in this questionnaire, agreeing to the written phrases below.

Thank you for your collaboration!

Questionnaire

1 - I have read and understood the meaning of this study. I had the opportunity to put questions, if necessary, and collect their answers.

Results: 100% agreed.

2 - I understood that the data collected in this study will be used as mentioned above.
Results: 100% agreed.

3 - I authorize the use of video recordings collected during the session.

Results: 100% agreed.

4 - I authorize the dissemination of audiovisual data collected during the session on online platforms (for example: youtube, vimeo, etc.)

Results: 100% agreed.

5 - As described above, I authorize my participation in this study and accept its conditions.

Results: 100% agreed.

6 - I understand that the participation in this study is voluntary and that I can withdraw at any time without any explanation. If this happens, I will not be subject to any penalty and my experience data will be removed and destroyed.

Results: 100% agreed.

7 - I authorize the processing of data within the framework of this project for the purpose of analysis, research, and dissemination of results in scientific publications or conferences in the project area by the researchers of this project.

Results: 100% agreed.

User Profile

![Gender Profile](image-url)
Appendix B BioSignalsPlux Experimental Work

BioSignalsPlux

In this study, it was used two components of BioSignalPlux Researcher. The wireless 8 channel hub that receives the data from the sensor by wireless, and the piezoelectric respiration sensor for respiratory analysis. In Figure 33 can be seen the both components used.

The hub has 8 analog inputs (16-bit per each channel), the sampling rate is up to 3000 Hz (per channel), and the communication use Bluetooth 2.0+EDR, and the range of this communication is around 10 m. This device weight 45 g and during the acquisitions can be stuck in the belt of the sensor.

![Wireless 8 channel hub; Piezoelectric respiration sensor.](image)

**Piezoelectric Sensor**

The Piezoelectric Sensor uses the piezoelectric effect to measure changes in pressure, acceleration and force, that are converted into electric voltage:

\[
PZT(\%) = \left(\frac{ADC}{2^n} - \frac{1}{2}\right) \cdot 100\%
\]

Where \(PZT(\%)\) is the displacement value in percentage of full scale, the \(ADC\) is the value sampled from the channel and \(n\) the number of bits of the channel. The magnitude range of the output signal are between -50% and 50%.


Preliminary Experimental work

In order to test the reliability of Kinect skeleton points to measure the respiratory rate, it was decided to compare the data from the Kinect skeleton with the data from BioSignalsPlux. The data from the BioSignalsPlux was acquired and saved with the OpenSignals software, and the data from Kinect v2 was acquired and saved with a MATLAB code developed in MATLAB R2016a. Both software programs ran on a PC with an Intel(R) Core(TM) i7 CPU 4710HQ @ 2.50 GHz processor with 12 GB of RAM. The acquisition of the data was not performed synchrony.

To better analyze the performance of these points on measuring the breathing rate, it was decided to make four acquisitions for each participant, two in rest mode and two after-exercise modes. In addition, for each mode it was made half acquisition seated and the other half standing. All four acquisitions were performed for all the participants and lasted one-minute each mode.

The initial step was to recruit volunteers, to make the acquisitions. The volunteers recruited were all students. The experimental data was collected from 6 healthy subjects (4 males, 2 females; 24.66±1.03 years old; height 173.17±11.67 cm; weight of 68,17±10.89). Before the acquisitions started, one at a time, the participants responded to informed consent where they all understood the meaning of the study and authorized the recording, collection and processing of data.

In order to avoid failures, a guide was followed for this study. First, it was explained the study, the objectives and the procedures. Then the participants read and responded to the informed consent. At that point, it was made the firsts acquisitions, one-minute seated acquisition followed by one-minute standing acquisition. The next step was to tire the participant to test the system with high breathing rates. To increase the breathing rate, the volunteers climbed three floors of the tower quickly. This physical exercise was done before the last two acquisitions. These last two acquisitions were made in the same way that the first two, the only difference was that before each acquisition the participant made the physical exercise. During the acquisitions, it was collected data from Kinect and BioSignalPlux at the same time but not synchronously.

All participants were positioned at the same distance from Kinect sensor. The Kinect was placed at the railing of the window, it means 91 cm from the floor. On the seated tests, the chair was aligned with the z axis of the Kinect sensor and it was placed 140 cm from the sensor. On the standing tests, as well as the seated tests, the participant was aligned with Kinect and it also was placed 140 cm from the sensor.

To be able to get as much information as possible from these fifteen signals, it was tested an approach to get the best respiratory rate signal possible. The first approach was to calculate the root mean square of the axes of each point. Seeing the results of this approach, it has been proven that this is not a good option. This would only work if, for example, at an expiration, the sum of the squared axes was always positive or negative and the opposite in inspiration. For that reason, it was discarded.
In this experimental data was only analyzed the fifteen signals mentioned above and compared them with the signal from BioSignalsPlux. This comparison was done after applying the passband filter in both signals. To compute the frequency of each signal, the signals was not synchronized, it was count the peaks of each signal and divided by the time passed between the first peak and the last.

Results of the Preliminary Experiment

The results of the acquisitions in the preliminary experiment can be seen in Figure 34. In this figure, it can be seen the mean and the standard deviation of the error between the frequency of each signal from Kinect and the signal from BioSignalsPlux (ground truth). In a first approach, this graphic can show the most reliable axes.

Figure 34 - Mean and Standard Deviation in the method of points axes in the four modes of acquisition using the BioSignalsPlux. (Top left) Seated acquisition before exercise; (Top right) Standing acquisition before exercise; (Bottom left) Seated acquisition after exercise; (Bottom right)

The Figure 34 shows that is possible measure the breathing rate using the skeletal joints from Microsoft Kinect, mainly from the results of the seated acquisitions. Analyzing these graphics, it can be observed a pattern: for seated acquisitions the z-axes obtained less error and for standing acquisitions, the y-axes were the axes with a smaller error. It can also be concluded that the x-axes were, in general,
the worst axes, which makes sense since the respiration movements in these axes are not as relevant as in the other directions.

In the standing acquisitions, the error was excessively large, this issue can be justified by two factors: the subjects in standing acquisitions can oscillate giving rise to an error of measurement and it can be due to the fact that in standing acquisitions the movement is predominantly along the y-axis, which may provoke bad detection of the piezoelectric sensor. This detection problem of the piezoelectric sensor triggered the change to a new ground truth device.

To analyze the results from this experimental work, it was created the Table 2 where it can be seen the lower error value for each acquisition. This table has formed a way to check if it is practicable to use the skeletal points from Kinect to measure the respiratory rate.

### Table 2 - Error of the best axis in each acquisition

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>SITTING NORMAL BREATH</th>
<th>STANDING NORMAL BREATH</th>
<th>SITTING HEAVY BREATH</th>
<th>STANDING HEAVY BREATH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.181 bpm</td>
<td>3.199* bpm</td>
<td>0.050* bpm</td>
<td>0.222 bpm</td>
</tr>
<tr>
<td>2</td>
<td>0.037 bpm</td>
<td>0.092 bpm</td>
<td>0.019 bpm</td>
<td>8.661* bpm</td>
</tr>
<tr>
<td>3</td>
<td>0.633 bpm</td>
<td>3.618 bpm</td>
<td>1.104 bpm</td>
<td>3.872 bpm</td>
</tr>
<tr>
<td>4</td>
<td>0.008 bpm</td>
<td>0.160 bpm</td>
<td>0.543 bpm</td>
<td>5.881 bpm</td>
</tr>
<tr>
<td>5</td>
<td>0.133 bpm</td>
<td>7.174* bpm</td>
<td>1.547 bpm</td>
<td>0.069 bpm</td>
</tr>
<tr>
<td>6</td>
<td>0.766* bpm</td>
<td>0.036 bpm</td>
<td>0.137 bpm</td>
<td>0.333 bpm</td>
</tr>
</tbody>
</table>

In the Table 2 the values marked with ‘*’ are the results which the data acquisition from BioSignalsPlux was not able to capture the breathing movement correctly. These results show the high possibilities of this system to measure the breathing cycle. In the Figure 35 it can be seen an example of one acquisition that had this problem and one without it.
Figure 35 – (Left) Example of a plots of a bad acquisition, before and after filtering the signal; (Right) Example of a plots of a good acquisition, before and after filtering the signal.

Through the Table 2, it is visible that the Kinect signals can have reach values of breathing rate quite close to the values from BioSignalsPlux. The fact that BioSignalsPlux failed to correctly capture some acquisitions justifies the use of another device as ground truth. It was noticeable that in cases where BioSignalsPlux could not correctly detect the respiratory movement, some Kinect signals had a sinusoidal wave (before applying the filter) that may result in a close breathing rate of a subject.
Appendix C EVM approach

The starting point for the analysis of the problem given was the Eulerian Video Magnification algorithm [2]. This method was used to extract the respiratory rate through the points of the skeleton, until a final analysis method was developed.

As explained in the Literature Review chapter, this algorithm was made to magnify subtle movements in a video sequence. With this objective in mind, the MATLAB code provided in http://people.csail.mit.edu/mrub/evm/ was used. Since the algorithm required a video sequence as its input, it was necessary to convert the signal of the axis of the points from Kinect skeleton to a grayscale video sequence. The conversion was made creating a video sequence for each axis, where the values of the joint’s position were converted to greyscale.

Using the video input, it was possible to test the EVM algorithm directly. For our problem, it was chosen the amplify.spatial.Gdown_temporal.ideal function which contains a temporal filter that is an ideal passband and a spatial filter that does Gaussian blur and down sample. Apart from the input video, it was necessary to set some variables: the amplifying factor (set to 200), the sampling rate (set to 30), the lower and upper frequency bounds for the ideal temporal filter window (set between 5/60 Hz and 40/60 Hz), the level that indicates the band of Gaussian Pyramid for the spatial filter (set to 1), and the attenuation of the color in the magnification process (set to 1).

Using the returned output video, it was possible to detect the respiratory movements. In Figure 36, it can be seen a representation of a returned video of the EVM algorithm applied to the skeletal joints signals, along the time. However, the data size was unaffordable to aspire to do it in real-time. In order to solve this problem, a temporal filter that can be better used with an array rather than a video sequence is needed. For that reason, this approach was discarded.

Figure 36 - Example of the variations of a signal from skeletal joints after filtering and magnifying it (in 30 s).