DEVELOPMENTS IN FIBRE OPTIC CARDIAC AND RESPIRATORY PLETHYSMOGRAPHY

Francois-Xavier Maletras

A thesis submitted in partial fulfilment of the requirements of Kingston University for the degree of Doctor of Philosophy

October 2002

The following Figures and Publications have been omitted on request of the University –

Pour mes parents

Acknowledgements

Throughout the last four years, the presence of some individuals has had constructive and decisive influences on the outcomes of this research program.

I first would like to express my gratitude to my supervisors, Dr. Andy Augousti and Dr. Julian Mason, for their help and scientific guidance during the entire four years of this research program. The contact of Andy and Julian was always food for thoughts and our numerous technical and scientific discussions were exciting moments that I will not forget.

The technical expertise of Mr. David Higgs has had an important impact on this research. Dave not only designed all the necessary printed circuit boards but also gave me access to the large resources of his laboratory, permitting the prolonged use of excellent pieces of equipment.

No results would have been obtained without the kindness of the numerous people who accepted to take part in my sometimes repeatedly uncomfortable experiments. I am eternally obliged to their benevolence.

I am also perpetually indebted to Dr. Ralph Manly, Head of the School of Life Sciences, and Dr. Demetra Katsifli, Head of Academic Information Systems and Customer Services, ICT Services, whose sustained supports made this research project come into existence and reach its goals.

Demetra gave me the opportunity to work part-time as an IT helpdesk analyst, therefore partially funding my research, whereas Ralph was responsible for covering the entire cost of my tuition fees over the last four years.

My kind regards go to Dr. Christian Bugge "Bug" for being such a valuable friend in general, and more specifically in the context of our parallel experiences of postgraduate education at Kingston.

Finally, I would like to thank with all my heart Miss Natalie Riches for being such a great, special person, for her unshakable Love and inalterable Trust in me, and for her continuous support since last year.

Abstract

This work is the continuation of previous research by A. Raza and other contributors to the field of fibre optic plethysmography. Plethysmography is defined as the volume estimation of an object according to its external dimensions. In particular, this technique can be used to produce an estimation of the respiratory volume of a human subject according to the dimension of his chest, measured at the thoracic and abdominal levels.

A respiratory plethysmograph simply attempts to deliver a signal being the closest possible estimation of the true respiratory volume, as measured by a spirometer or a pneumotachometer. There are essentially two instrumental approaches to respiratory plethysmography:
1) The Respiratory Inductive Plethysmograph (RIP) estimates the cross section area of the chest by monitoring the variation of inductance in an electrical wire encircling the chest.
2) The Fibre Optic Respiratory Plethysmograph (FORP) sees the contribution of fibre optic sensors to measure the chest's circumference variations.

The purposes of the present investigation were to improve the performance of previous FORP prototypes, and to extend its capabilities to cardiac monitoring. Both these targets have been reached, and the new prototype is now demonstrating the potential of plethysmography as a sound investigation technique for both cardiac and respiratory monitoring.

Overall, the improved sensor and acquisition system permitted the resolution of details of the plethysmographic waveforms that were beyond the reach of the previous prototype. The new FORP prototype is generally more reliable and more precise, if not less compact. From a medical point of view, research carried out with the new FORP prototype has had two major outcomes:

1) The increased temporal resolution of the new acquisition system has given us the possibility to precisely measure the phase shifts between the plethysmographic signals, and the spirometric signal. Such measurements have contributed to producing a better estimation of the spirometric signal, therefore increasing the credibility of the FORP as a non-invasive, respiratory volume monitoring device.

2) The increased amplitude resolution of the new acquisition system, coupled with the better linearity, better precision and smaller hysteresis of the new sensor, has enabled the FORP to detect body circumference variations due to cardiac activity around head, neck, thorax and abdomen of a patient.

Observations of heart movements at thoracic level had already been reported with the RIP, the direct analogue of the FORP. The signal processing required by the RIP for such monitoring only permitted offline, Electro-Cardio-Gram (ECG) assisted interrogation of cardiac displacements.

However, thanks to better signal processing, the FORP has been made capable of real time cardiac position monitoring, without referencing to a simultaneous ECG signal.

The combined impact of this research and previous research by A. Raza and A. Augousti on respiratory gating with the FORP, is potentially important in the field of cardiac imaging with Magnetic Resonance and Computed Tomography scanners. The FORP should permit better synchronisation with cardiac movements, while helping the patient to maintain stable chest position, subsequently increasing the image resolution by limiting motion blur.

List of Abbreviations

AV Atrioventricular

C(x,y) Covariance coefficient

CMR Common Mode Rejection

CT Computed Tomography

DCP Digitally Controlled Potentiometer

DFT Discreet Fourier Transform

DSP Digital Signal Processing

EBCT Electron Beam Computed Tomography

EM Electro-magnetic

FFT Fast Fourier Transform

FOP Fibre Optic Plethysmograph

FORP Fibre Optic Respiratory Plethysmograph

FT Fourier Transform

GUI Graphical User Interface

HFOV High Frequency Oscillatory Ventilator

IC Impedance Cardiography

IFFT Inverse Fourier Transform

ILT Inverse Laplace Transform

IO Input Output

IP Impedance Pneumography

IR Infra Red

IR-LED Infra Red Light Emitting Diode

IRV Inspiratory Reserve Volume

ISA Industry Standard Architecture

ISR Interruption Subroutine

IZT Inverse Z Transform

LED Light Emitting Diode

LT Laplace Transform

MBLE Macro-Bending Loss Effect

MPD Mode Power Distribution

MR Magnetic Resonance

MRI Magnetic Resonance Imaging

NMR Nuclear Magnetic Resonance

PSD Power Spectral Density

PSU Power Supply Unit

RAM Random Access Memory

RF Radio Frequency

ROM Read Only Memory

RV Residual Volume

SH Sample & Hold

SNR Signal to Noise Ratio

TIDA Transimpedance Differential Amplifier

VC Vital capacity

VCA Voltage Controlled Amplifier

VCO Voltage Controlled Oscillator

VM Virtual Machine

ZT Z Transform

Table of Contents

Chapter 1 INTRODUCTION

In biomedical research, as well as in clinical medicine, the acquisition of physiological data from the patient is important not only to the understanding of biological mechanisms but also to support diagnostic procedures. The quality of such measurements largely depends on the performance of the transducers and the acquisition system. The living body is an uneasy terrain for accurate measurements because it requires transducers capable of isolating a particular physiological phenomenon, without interfering with this phenomenon or other simultaneous phenomena.

Many measurement challenges in this field have been overcome thanks to the variety of sensors that has been developed.

In respiratory monitoring for example, at least five different techniques can be enumerated, including the one presented in this thesis.

When it is possible, non-invasive measurement techniques are always preferred as they guarantee minimal interference with the body's functioning, by respecting the natural skin barrier.

The present investigation in biomedical engineering aims at presenting improvements in the field of fibre optic plethysmography applied to non-invasive cardiac and respiratory monitoring.

Augousti and Raza originally developed the Fibre Optic Respiratory Plethysmograph (FORP) at Kingston University, UK, between 1992 and 1998 [Augousti and Raza (1993), Raza and Augousti (1994), Raza et al. (1995), Augousti et al. (1996), Augousti (1997)]. This instrument was initially designed for assessment of breathing patterns but other researchers in the field of plethysmographic instrumentation have since demonstrated the potential of plethysmography for cardiac monitoring.

The starting point of the present investigation, in October 1998, was the continuation of the work by Raza and Augousti. Outcomes of their original research have been amalgamated with findings from other contributors to plethysmography, in order to create novel modalities of cardiac and respiratory monitoring with the FORP.

The new research results presented in this thesis are based on plethysmography, optical fibre instrumentation and Digital Signal Processing (DSP). The implications of this research are for cardiac and respiratory monitoring, as well as medical imaging.

1.1. Respiratory monitoring

The most obvious use for respiratory monitoring is in the clinical diagnosis and prognosis of respiratory defects. When detected, the type of pulmonary defect of a condition can be classified by a characteristic breathing pattern and often, such as in the case of the spirometer, with enough accuracy to detect the abnormalities in breathing before the onset of serious illness.

The spirometer and pneumotachometer (further detailed in section 2.5.1) are the only two investigation methods for absolute volumetric measurements of respiration, i.e. they permit the immediate measurements of different respiratory volumes. These methods are invasive and, in the case of the spirometer, non-ambulatory.

These instruments also frustrate the need for continuous respiratory monitoring (during sleep for example) in that their invasiveness is a major source of discomfort for the patient.

The spirometer and pneumotachometer require the subject to breathe through a mouthpiece, which is sometimes supported by a facemask. The discomfort of wearing a nose clip and a mouthpiece, coupled with having to breathe against a backpressure from the instrument, makes prolonged use of the spirometer or pneumotachometer impractical. This hindrance is even more apparent in frail and young users.

Both instruments contain numerous mechanical parts exposed to gaseous flow exchange with the subject's airways. In addition, the spirometer requires adequate oxygen supplies and constant removal of carbon dioxide for any investigation longer than one minute. Consequently, very frequent maintenance is required to disinfect each instrument and ensure it is correctly functioning.

In contrast, non-invasive plethysmographic respiratory monitoring, because it does not require permanent contact with respiratory airways, allows continuous investigation by strongly limiting the discomfort of wearing the instrument.

1.2. Cardiac monitoring

Heartbeats are the first vital sign to be checked by medical practitioners when encountering patients of unknown health conditions. This underlines the importance of the heart's physical condition to sustained life.

The most common non-invasive cardiac diagnostic and prognosis procedure is by analysis of the patient's ElectroCardioGram (ECG), i.e. the interpretation of the patterns of the heart's bioelectric signals, measured at the surface of the skin. Basic ECG recording techniques using three electrodes

("three lead ECG") are relatively simple. In more advanced techniques involving twelve electrodes ("twelve lead ECG"), thorough assessment of the heart's condition can be achieved.

Several million diagnostic ECGs are recorded every day. The ECG is also continuously monitored in intensive care in that it provides irreplaceable information about the patient's status, in real time.

Given that cardiovascular diseases are among the first causes of death in the occidental world, frequent investigation of heart condition is important for people at risk. This is becoming increasingly important as the average age of the occidental population continues to rise.

In these conditions, it is easy to understand why commercialisation of ECG systems is an important and lucrative business.

1.3. Cardiac imaging

Other medical tools for diagnostics of cardiac and/or cardiovascular diseases are ultrasound echography, Magnetic Resonance (MR) imaging and the X-ray Computed Tomography (CT). These systems permit the visual representation of the heart.

Despite its high temporal resolution (multiple images per second), echography has a lower spatial resolution than the MR and CT scanners. These last two imaging systems have a lower temporal resolution but permit accurate visual representation of the heart (cardiography) and the coronary system (angiography). One of the interests in angiography is the localisation and measurement of plaques formed by calcification of the arteries.

To perform cardiac or cardiovascular imaging, both MR and CT scanners need to be synchronised to the patient's cardiac cycle, in order to limit motion blur due to heart or blood movements. Synchronisation is usually provided with an ECG system. While the ECG does not interfere with the functioning of the CT scanner, this is not the case in the MR scanner: Due to the harsh electromagnetic environment generated by the MR scanner, the presence of metallic components (the ECG electrodes and leads) in the immediate vicinity of the patient presents an immediate health and safety issue.

1.4. Fibre optic plethysmography for cardiac and respiratory monitoring

The concept of non-invasive respiratory monitoring initially originated from the reported observations that invasive respiratory monitoring instruments such as the spirometer and pneumotachometer hinder breathing efforts [Gilbert et al (1972), Dolfin et al (1983), Maxwell et al (1985), Paek and McCool (1992)]. The idea of a non-invasive respiratory monitoring instrument leaving the mouth and nose unobstructed therefore has an immediate appeal. This point, and the historical development of plethysmographic respiratory monitoring, have been thoroughly discussed by A. Raza [Raza (1998)].

Fibre optic plethysmography is a numerously reported non-invasive cardiac and respiratory monitoring technique, based on the measurement of chest circumference variation at thoracic and abdominal levels with two fibre optic elongation sensors.

Plethysmography, by itself, is the process of estimating the volume of an object from the measurement of its external dimensions.

This principle can be applied to human physiology, since both cardiac and respiratory activities contribute to the production of torso volume variations, and that subsequent measurements of torso circumference variations can be utilised for estimation of volumetric variations of the heart and/or the lungs. Cardiac monitoring is possible during ventilation under certain conditions but is greatly facilitated by interruption of breathing.

The plethysmographic instrument discussed in this thesis is the Fibre Optic Respiratory Plethysmograph (FORP). The optical fibre sensor used in the FORP is arranged in a specific configuration aimed at encouraging the Macro-Bending Loss Effect (MBLE) when stretched. At the sensor's input is an optical signal of stable power. The MBLE modulates this input power according to the size of the extension applied to the sensor. Subsequently, this extension can be quantified by measuring the optical power at the sensor output.
The sensor can be used to measure elongation along a single axis or around an elliptical area like the cross section of a human torso.

1.5. Fibre optic sensors

The second half of the 20th century was marked by a dramatic advance in the capacity of transmission of information. This was partially due to the enormous improvement of optical fibre communication systems.

In 1970, a new generation of optical fibre with an unprecedented low power attenuation of 20 dB/km, at the He-Ne laser wavelength of 633 nm, became available and made optical fibre communication systems practical. Since then, both light sources and optical fibres have undergone continuous development and power attenuation of the order of 0.1 dB/km is now common for most communication systems using monomode optical fibre and infra-red laser diodes.

Optical fibres now transmit and receive data over much longer distances than copper wires, for identical signal attenuation. The density of information that can be propagated by optical fibres is also higher than for copper cables.

The information propagating through the optical fibre system is carried by a luminous signal. The information source is coded in the signal by modulating the power allowed into the light source. Once modulated, the light signal is conveyed by the optical fibre and demodulated by the receiver on the other side of the optical cable. The optical fibre consists of a cylindrical dielectric core of high refractive index surrounded by a lower refractive index cylindrical cladding. The core is usually based on silica materials. Because of the difference between core and cladding refractive indices, the light is guided inside the fibre core by the mechanism of Total Internal Reflection (TIR). The optical power dissipated by transmission of light from the core to the cladding is insignificant compared to the optical power carried by the fibre core.

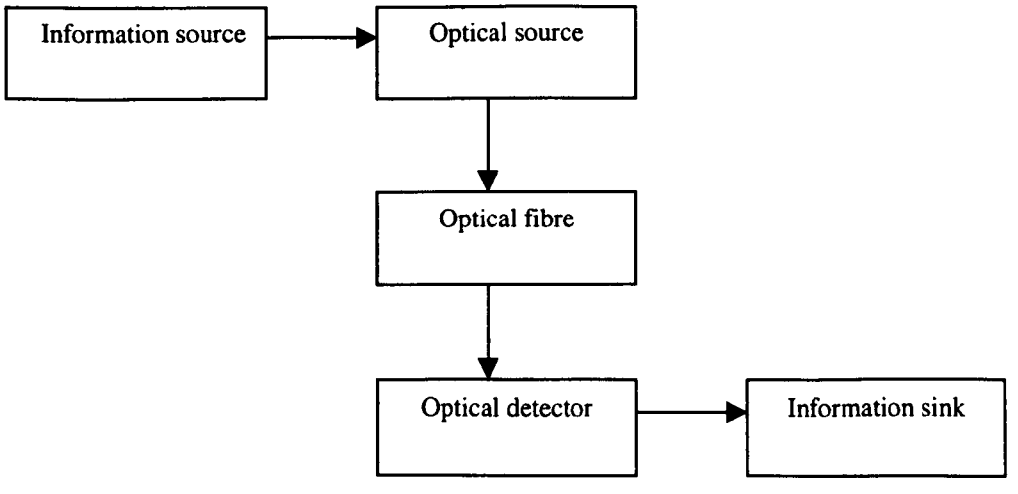

Figure 1. The optical fibre communication system

1.5.1. Emergence of Optical Fibre Instrumentation.

Progress in optical fibre instrumentation has closely followed the recent advance in optical fibre communication, as they both depend on optical fibre manufacturing technology. Optical fibre sensors have many advantages over metallic electronic sensors.

Optical Fibre based measurement (OFBM) concerns an increasing list of different physical quantities, including rotation, acceleration, electric and magnetic field measurement, temperature, pressure, acoustics, vibration, linear and angular position, strain, humidity, viscosity and chemical measurements, to name but a few. Because of this large field of applications, OFBM systems have rapidly expanded and optical fibre systems were designed not only to innovate but also to compete with already existing electronic or mechanical instruments. This makes OFBM an equivalent if not better approach to measurement problems in many fields of instrumentation. Optical fibres are formed of dielectric waveguides, which makes them electrical insulators. In consequence, they are relatively safer to manipulate than electrical wires and have complete immunity to electro-magnetic interference (EMI). EMI immunity also guarantees that the transmitted signal is insensitive to cross talk if there are other optical fibres in the vicinity. Additionally and perhaps more importantly, the decreasing manufacturing cost of optical fibre and optical fibre related components, typically semiconductor light sources, coupled with a constant improvement in component quality has encouraged instrument manufacturers to favour the optoelectronic approach to instrument design problem solving. When the laser diode (for example) became commercially affordable and reliable in the early 1990, it rapidly became popular with its use in household applications such as portable compact disc players and laser printers.

An optical fibre instrumentation system is similar in principle to an optical fibre communication system with the difference that the optical fibre instrument is designed to let the propagation of light in the optical fibre deliberately be disturbed by a physical phenomenon. The physical quantity of interest is measured according to this disturbance in light propagation. Modulation of the light source input power has no more role in information coding than as a measurement parameter. For example, in the optical fibre instrument that will be presented in this thesis, the light source input power is pulsed to achieve larger optical output power than is normally possible in continuous mode but this has no effect on the physical quantity being measured.

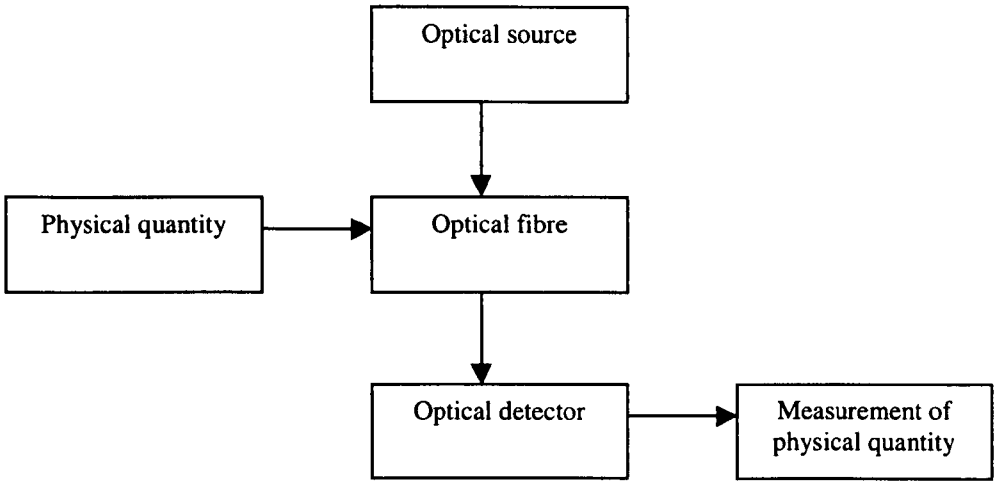

Figure 2. The optical fibre instrumentation system

1.6. Digital Instrumentation and Signal Processing

To sum up the evolution of electronic engineering in the last fifty years, one might say that the semiconductor revolution we live in started with the invention of transistors. The transistor has greatly benefited analog electronics (essentially by replacing the valve tube). Due to the restricted utilisation of analog signals to only two voltage levels, it has also permitted the easy implementation of Boolean algebra in electronic systems. This particular area of electronics, using the logical rules of Boolean algebra and binary signals to transport logical information, was termed digital electronics.

Digital electronics permitted the construction of transistor based, programmable binary computers, achieving an enormous size reduction and speed increase compared to previous non-transistor computers (either analog or digital).

Since then, the computational power of these digital machines has never ceased to increase, hence permitting an ever-larger scope of applications, and in particular in the field of instrumentation. This is possible for two reasons. Firstly, any analogue system operating under the speed limit of digital technology (approximately 10^{10} operations per second nowadays) can be simulated in a digital manner. Secondly, digital technology feeds its own development by becoming ever cheaper, because of mass production.

As a result, the tendency of electronic engineering (including instrumentation) is to become 'all digital'. Multimeters, spectrum analysers and oscilloscopes are only a few obvious examples of mass-produced digital measurement instruments originally developed as analog systems.

The massive progression of digital technology is also based on its ability to efficiently circulate information. The information is universally carried by a signal, either analog or digital. In order to encode or decode information the signal must be processed.

The fields of communication and instrumentation have had an incomparable impact in the development of modern Digital Signal Processing (DSP). The two first domains share the same goal: coding and/or extraction of information from signals. DSP regroups the ever-expanding number of logical and numerical techniques that can be applied to a signal (either digital or analog) to retrieve its information.

Naturally DSP also allows simulating Analog Signal Processing (ASP) but the contrary is not necessarily true. For example, DSP permits the implementation of both Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) filters, whereas ASP filters are IIR by nature.

ASP is still widely used but is progressively retracting, there are cases however when DSP cannot supersede ASP. For example, the issue of analog to digital conversion lies exactly at the frontier of analog and digital electronics. In this particular case, both techniques are necessary because the natural area of the expression of physical quantities is primarily the analog domain.
The use of DSP only becomes possible after the signal's discretisation: the analog signal is first sampled in time and the values of the resulting samples are then quantified with a binary code.

When measuring physical quantities, it is important to employ instruments with stable characteristics. By manipulating signals in their binary form, DSP permits a much better permanence of the instrument specifications than ASP. Digital electronics offers neither short-term nor long-term drift with temperature or supply voltage fluctuations, and therefore achieves a high reproducibility. The accuracy of a digital instrument is determined by 1) the number of bits used in signal quantisation, 2) the rate of acquisition and 3) the accuracy of the analog sensor attached to the digital acquisition and processing system.

Spectral analysis, filtering, and modulation are only a few of the most popular DSP operations. DSP applies the versatility of programmable digital electronics to the never-ending quest for better communication and better understanding of ourselves and the world we live in, through accurate measurements of physical signals.

1.6.1. Signal Processing in Medical Instrumentation

The field of medical instrumentation is concerned with understanding physiological phenomena by interpreting measurements of diverse biomedical signals.

Medical instrumentation greatly benefits from advances in DSP for the reason that raw biomedical signals very often require a large number of processes in order to reveal their meaning. For example, dedicated filtering techniques are being developed for the sole purposes of extracting the ElectroCardioGram (ECG) of an unborn child from his mother's ECG. This particular research problem is difficult because the mother's ECG is on average 500% larger than her child's, while both ECG traces share the same frequency range. Once cleared up, a spectral analysis of the child's ECG can permit pre-natal detection of cardiac anomalies. This example illustrates the need for specialised signal processing in the field of medical instrumentation.

Biomedical signals belong to a group of signals with specific statistical properties. Generally speaking, such signals have low frequencies, low signal to noise ratio and are prone to instrumental or physiological artefacts. Table 1 below summarises the statistical characteristics of typical human biomedical signals.

ElectroCardioGram	Average rate: 45 to 200 BPM Frequency range: 0.05 Hz to 100 Hz. Signal range: 0.01 mV (fetal) to 5 mV (adult)
Blood pressure	Pressure range: 40 to 300 mmHg (arterial), 0 to 15 mmHg (venous)
Respiration	Average rate: 12 to 40 BPM Nomal flow range: 250 to 500 ml/s. Tidal volume: 600 ml/breath (adult)
ElectroEncephaloGram	Frequency range: DC to 100 Hz. Signal range: 0.015 mV to 0.1 mV

Table 1: Statistical characteristics of some human biomedical signals. We can see that the frequency range of these signals never exceeds 100 Hz, which is rather small. Therefore, a fast acquisition rate is not a priority in designing a biomedical instrument, however, given the large range of amplitude variations, it must be precise and accurate.

Without exception plethysmography has benefited from the emergence of DSP. It is interesting to note that both fields came into existence during the 50's and have progressed in parallel. Respiratory plethysmographic signals generated by the early system from Konno and Mead were recorded on paper with an analog plotter and later analysed manually to produce the necessary calibration constants.

In contrast, by 1998 Raza implemented multiple simultaneous real time calibration methods using a laptop computer, resulting in multiple sets of calibration coefficients to be automatically determined.

Apart from the laptop computer and a miniature signal acquisition module for the mentioned computer, the system consisted of a "home-made" signals amplification box and two prototype plethysmographic sensors. In designing the respiratory monitoring system, the most important part of the software work was with coding of the statistical DSP methods for generating the calibration coefficients.

1.7. Thesis overview

The present thesis is mainly divided into three parts:

1) Chapters 2 to 6 cover background materials in the fields of plethysmography, optics, signal processing, physiology and medical imaging. It is intended that the information provided here is sufficient for the purpose of understanding the principles and applications of the FORP.

2) Chapter 7 and 8 present a thorough review and discussion of the principles and modalities of cardio-respiratory monitoring with plethysmography.

3) Chapters 9 to 12 concern the outcomes of the present research. These cover essentially the areas of plethysmographic techniques (chapters 9 and 10), the refinement of plethysmographic respiratory monitoring (chapter 11) and the new development of cardiac monitoring (chapter 12).

In chapter 2, the structure of the human respiratory system is presented and its mechanical functioning is described at a physical level. Also, all non-plethysmographic respiratory monitoring methods are briefly surveyed. This provides the grounds of a comparison between respiratory plethysmographic monitoring and the concurrent techniques.

In chapter 3, the heart and cardiovascular system are described in physiological terms, and the source of the ECG signal is explained. This allows an understanding of the correlation between the mechanical and electrical activity of the heart.

In chapter 4, elements of optical physics relating to the propagation of light in optical fibres are covered. The Macro-bending Loss Effect (MBLE), in use with the FORP sensor, is also studied. A comparison of different MBLE theories is provided.

Chapter 5 contains elements of Digital Signal Processing theory required for the understanding of the research presented in chapters 11 and 12.

Chapter 6 is a brief description of the principle of the MR scanner. This enables us to understand the specific conditions of operation of the MR scanner, and how the FORP is compliant with these conditions.

Chapter 7 concerns plethysmographic respiratory monitoring. This covers plethysmographic respiratory models, their methods of calibration and the different respiratory plethysmographic instruments, including the FORP. The principles of respiratory plethysmography are reviewed and discussed in details.

In chapter 8, the extension of plethysmography to cardiac monitoring is discussed and the range of cardiac plethysmographic instruments (including the FORP) is presented.

In chapter 9, a new sensor design for the FORP is reviewed. The original FORP sensor has been modified to contain a single coil of optical fibre (as opposed to the multiple coil design previously used). The new sensor design, termed the figure-of-eight coil, has the advantage of providing an enhanced resolution and a recoil force that naturally reduces the hysteresis. The new sensor is simpler to construct and to operate, also offering better performance in terms of linearity, elongation range, reproducibility and robustness. Proof is given that its characteristics are superior to other FORP sensors presented in chapter 8.

In chapter 10, the issue of designing a new acquisition system was addressed. The newly developed system is PC based, and offers an extremely good improvement of signal quality compared to the first version of the system, as developed by A. Raza during his research.

Chapter 11 presents the details of a new calibration technique for respiratory monitoring. This new method is based on the independent correction of the gain and phase of the thoracic and abdominal signals. Results obtained with the new method are compared with a previous method. This calibration method produces estimations of ventilatory signals that are more closely related to the real absolute ventilatory signals than any other calibration method.

In chapter 12, proofs are given that cardiac activity has been detected with the FORP at abdominal, thoracic, carotidal and cranial levels on human subjects. The principle of precise cardiac period measurement is explained in detail. Such measurement was made possible thanks to specifically developed signal processing algorithms.
It is suggested that, because the thoracic cardiac activity is a reflection of the heart ventricular volume, the FORP could be used to trigger a cardiovascular imaging system such as an MR scanner or a CT scanner.

Chapter 13 contains a discussion of the work and its implications and a conclusion. Further work necessary to extend the present investigation is also proposed.

Chapter 2 RESPIRATION AND RESPIRATORY MONITORING

Part of the research work presented in this thesis is concerned with respiratory monitoring. The elements of human physiology given in this section are aimed at providing the necessary knowledge of the respiratory system for understanding the interpretation of the FORP signals.

2.1. Organisation of the respiratory system

Respiration serves the purpose of oxygen provision to the blood and elimination of carbon dioxide from the blood. This is realised within the lungs at the alveolar level, in a process of gas exchange between the blood and the air. Ventilation is the circulation of air in and out of the lungs. Inspiration is the movement of air from the external environment through the airways into the lungs during breathing. Expiration is the movement of air in the opposite direction. An inspiration followed by an expiration constitute the respiratory cycle. In humans and other mammals, the respiratory system includes the lungs, the series of tubes leading to the lungs, and the chest structures responsible for moving air into and out of the lungs during breathing. The lungs are the principal organs of the respiratory system. They fit inside the ribcage that protect them while allowing their displacement. During respiration, air passes through either the nose or the mouth into the pharynx, a passage common to both air and food. The pharynx branches into two tubes, the oesophagus, through which food passes into the stomach, and the larynx, which is part is part of the airways. The larynx opens into a long tube, the trachea, which in turns branches into the two bronchi, one of which enters each lung. Within the lungs, there are more than twenty generations of branching, each resulting in narrower, shorter and more numerous tubes. The smallest level of branching is terminated by the alveoli where gas exchange between air and blood takes place. The total alveolar gas exchange surface area represents 60 to 80 m^2 for an adult human. The walls of the trachea and bronchi contain cartilage, which gives them their cylindrical shape and supports them. The airways are surrounded by smooth muscles, the contraction and relaxation of which can alter airway radius. During the entire respiratory cycle, the right ventricle of the heart pumps blood through the capillaries surrounding each alveolus. At rest, in a normal adult, approximately 4 litres of fresh air enters and leaves the alveoli per minute, while 5 litres of blood flows through the pulmonary capillaries. During heavy exercise, the air flow increases by a factor of 20 and the blood flow increases by a factor of 6.

Figure 1: Anatomy of the human ventilation system [Clancy et al (1995)].

2.2. Respiratory muscles

Lung inflation and depletion is achieved by changing the dimensions of the chest wall by means of the respiratory muscles, the principal respiratory muscles are the diaphragm and the internal and external intercostal muscles. In addition some other muscles which are not involved during exercise may be used. These are the accessory muscles which assist in expiration. The diaphragm is the principal muscle of respiration and, during quiet breathing, is normally the only active muscle of respiration. The diaphragm forms a continuous sheet that separates the thorax from the abdomen. At rest, it assumes a dome-like shape. When it contracts during inspiration, it flattens, so increasing the volume of the chest. In the absence of nervous activation, it passively relaxes and this forms expiration. When the demand for oxygen increases, the chest wall is lifted upward and outward by the activity of the external intercostal muscles and the diaphragm contracts more strongly. In severe exercise, the accessory muscles are called on to lift the chest wall further during inspiration and the internal intercostal muscles contract to assist in decreasing the volume of the chest. Powerful expiration may also be assisted by contraction of the abdominal muscles that force the abdominal contents against the diaphragm, pushing it upwards and reducing the volume of the chest.

Respiration is not a completely autonomous body function as respiratory muscular control can be regained at will by the conscious mind. This factor, added to the relative freedom of displacement of thorax and abdomen, contributes to make respiration a complex body movement with relatively little predictability.

2.3. Ventilation mechanics

The chest wall is lined by a membrane called the parietal pleura. This is separated from the visceral pleura, which covers the lung, by a thin layer of liquid that serves to lubricate the surfaces of the pleural membrane as they move during respiration. In other words, the lungs are separated from the chest wall only by the pleural membranes and, in health, they occupy almost all the cavity of the chest.

The pleural membranes form a closed space of constant volume called the pleura. Low pleural viscosity allows the reduction of friction between the lungs and chest wall during their respective displacement.

Lungs are passive elastic structures, they have an inherent elastic recoil due to their tendency to collapse on themselves with their own weight. The thoracic wall also has a natural elastic recoil but in the opposite direction to the lungs: in the absence of contradictory forces, the chest wall would naturally expand outwards. At rest, when respiration is interrupted, both lungs and chest wall exert opposite and balancing forces on each side of the pleura, which provides an effortless[1] equilibrium position of the ventilation system. Generally speaking, the pleura has an expanding action on the lungs and a compressing action on the chest wall. Because the pleura is stretched on both sides, the intrapleural relative pressure is – 4 mm Hg below atmospheric pressure. During inspiration, the expansion of the chest wall and the diaphragm enlarges the contact surfaces with the parietal pleura which results in lowering the intrapleural pressure because the pleural volume is constant. The decreased intrapleural pressure applies in return an expanding force on the visceral pleura, hence on the lungs. If we consider alveolar air to be a perfect gas, Boyle's law is enunciated as:

$$P_{alv}V_{alv} = nkT \qquad \text{(Eq.1)}$$

where P_{alv} is the alveolar pressure in Pascal (Pa) , V_{alv} is the lung volume in m^3, n is the density of gas molecules per unit volume in m^{-3}, k is Boltzmann's constant and T is the temperature in Kelvin.

Boyle's law allows us to state that, for a constant temperature T and a constant quantity of perfect gas n, the alveolar pressure is inversely related to the lung volume, hence the lung volume enlargement creates a depression.

---

[1] No muscular control is involved.

At the extremities of the airway, (the mouth and nose), the pressure is equal to the atmospheric pressure P_{atm}. According to the law of bulk flow, the decreased alveolar pressure P_{alv} will provoke air to flow from an area of higher pressure (outside the lungs) into the area of decreased pressure (inside the lungs). In our case, the law of bulk flow is enunciated as:

$$P_{alv} - P_{atm} = Q \cdot R \qquad \text{(Eq. 2)}$$

where P_{atm} is the atmospheric pressure at the extremities of the airways, P_{alv} is the alveolar pressure, Q is the volume flow rate[2] and R is the airflow resistance. As long as $P_{atm} > P_{alv}$, the flow Q remains positive and air fills the lungs. As the quantity of air inside the lungs increases, so does the alveolar pressure (Boyle's law). When P_{alv} reaches P_{atm}, the difference in atmospheric and alveolar pressures is null and consequently air stops flowing, marking the end of inspiration.

Passive expiration is a relaxation phenomenon where, in the absence of respiratory muscle drive, the chest wall expansion cannot be sustained and consequently, the ventilation system returns to its equilibrium position, pushing air out of the lungs. With the help of internal intercostal muscles, expiration can also become active (therefore faster and stronger).

2.3.1. Compliance

In ventilation mechanics, the concept of compliance is used to relate the lung volume variation to a given variation of pressure. The transpulmonary pressure (or lung recoil pressure) is defined as the difference between the alveolar pressure P_{alv} and the intrapleural pressure P_{ip}. Both intrapleural and alveolar pressures have effects on the lung volume. The difference of these two pressures corresponds to the net external lung pressure responsible for lung volume variation. The lung compliance is the ratio of the lung volume variation change to the transpulmonary pressure variation. Lungs compliance is defined below as:

$$C_L = \frac{\Delta V}{\Delta(P_{alv} - P_{ip})} \qquad \text{(Eq. 3)}$$

where C_L is the lung compliance, ΔV is the lungs volume variation and $\Delta(P_{alv} - P_{ip})$ is the corresponding transpulmonary pressure variation. The compliance is a measure of the ease with which chest volume can be changed. Therefore, the compliance relates to the mechanical resistance of the lungs.

[2] The volume flow is the derivative of the volume with respect to time.

2.3.1.1. Static compliance

Static compliance is measured in the absence of air movement into or out of the lungs (in which case the alveolar pressure is the atmospheric pressure at all times and consequently, the transpulmonary pressure variation simply becomes the intrapleural pressure variation). In this case, the compliance is constant i.e. static[3] and the lung volume variation is directly proportional to the transpulmonary pressure. This can be written as:

$$C_L = -\frac{\Delta V}{\Delta P_{ip}} \qquad \text{(Eq. 4)}$$

This tells us that, in the absence of airflow, the resistive mechanical forces associated with lungs volume variation are of an elastic nature. There are two major determinants of static lung compliance. One is the stretchability of the lung tissues, particularly their connective tissues, and the other is the surface tension at the air-water interfaces within the alveoli Certain alveolar cells secrete a detergent-like substance known as the pulmonary surfactant, which reduces the cohesive forces between water molecules on the alveolar surface.

Therefore, surfactant lowers the surface tension, which increases lung compliance.[4] A low static lung compliance means that to produce a given amount of lung volume expansion, a lower pressure than normal intrapleural pressure must be developed, this requires more vigorous contraction of the diaphragm and inspiratory intercostal muscles.

---

[3] Except when lungs volume is close to the total lung capacity (maximum) or to the residual volume (minimum).

4 A striking example of surfactant-deficient lungs is the disease known as respiratory distress syndrome in neonates. This is the second leading cause of death in premature infants, in whom the surfactant synthesising cells may be too immature to function adequately. Because of low lung compliance, the infant is able to inspire only by the most strenuous efforts, which may ultimately cause complete exhaustion, inability to breathe, lung collapse and death. Therapy in such cases is machine-assisted breathing with High Frequency Oscillatory Ventilation (HFOV) and the administration of natural or synthetic surfactant via the infant's trachea. The FORP has successfully been previously used to monitor infant artificial respiration in HFOV [Davis et al (2000)].

2.3.1.2. Dynamic compliance

If we were to measure the lung compliance during ventilation, it would no longer be static, therefore indicating that the airflow provokes a non linear variation of the lung volume as a function of the transpulmonary pressure variation. During ventilation, the alveolar pressure is constantly varying and this introduces a dependence on the airflow in the compliance equation, thus explaining the change in compliance. We recall that the variation in alveolar pressure during respiration is explained by the law of bulk flow and can be expressed as:

$$P_{alv} - P_{atm} = Q \cdot R \qquad \text{(Eq. 5)}$$

If we consider the variation in alveolar and atmospheric pressures, we can write:

$$\Delta(P_{alv} - P_{atm}) = \Delta P_{alv} = \Delta(QR) \qquad \text{(Eq. 6)}$$

Embedding the above expression into the compliance expression yields:

$$C_L = \frac{\Delta V}{\Delta(QR) - \Delta P_{ip}} \qquad \text{(Eq. 7)}$$

This last expression shows the contribution of the airflow to the compliance. During inspiration, the dynamic compliance is lowered by the airflow from the mouth to the lungs whereas during expiration, the compliance is increased by the airflow out of lungs. Consequently, a representation of the variation of the lungs volume as a function of the intrapleural pressure shows an hysteresis loop. The usefulness of the dynamic compliance in clinical condition is that analysing the trajectory of C_L along the loop with respect to time permits to diagnosis particular types of respiratory defects.

2.4. Standard lung volumes

When the chest is expanded to its fullest extent, the amount of air it contains is called the total lung capacity. After a maximal expiration, the remaining volume of air that cannot be expelled is called the Residual Volume (RV). The difference between the total lung capacity and the residual volume is the vital capacity (VC). The air inhaled and exhaled with each breath is known as the tidal volume V_T. The difference in volume at the end of a normal inspiration and the vital capacity is known as the Inspiratory Reserve Volume (IRV). The amount of air that can be forced out of the lungs after a normal exhalation is called Expiratory Reserve Volume (ERV). The Functional Residual Capacity (FRC) is the volume of air left in the lungs at the end of a normal expiration.

The tidal volume varies according to the requirements of the body for oxygen. Consequently, the inspiratory and expiratory reserve volumes are variable. In contrast, for a given individual, the vital capacity and residual volume are relatively fixed.

2.5. Respiration monitoring

Respiratory monitoring (or pneumography, or ventilometry) is the study of the volume of air that moves in and out of the chest during breathing cycles. The purpose of respiratory monitoring is to assess the lung volume variation against time under different physiological conditions such as rest or physical effort. Quantitative and qualitative knowledge of breathing patterns assisting determining specific lung pathologies. There are two classes of respiratory monitoring: 1) invasive respiratory monitoring is a direct volumetric measurement method. The subject is required to exchange air with a volumetric instrument through a facemask. 2) Non-invasive respiratory monitoring is not a true volumetric measurement method but a volume estimation method where the lung volume is calculated from the dimensions of the subject's chest. The FORP enters this second category.

2.5.1. Invasive monitoring

Invasive respiratory monitoring plays an important role in the research work presented here as such technique was used to provide a reliable reference for lung volume measurement for the purpose of calibrating and assessing the performance of the FORP. The reliability of invasive respiratory monitoring is due to direct volumetric measurement of the air exchanged between the lungs and the surroundings.

2.5.1.1. Spirometer

The spirometer is the classic respiratory monitoring device. It enables the direct measurement of all lung volumes, with the exception of the residual volume, as it cannot be expelled from the lungs. The spirometer is a pneumatic system. The model employed during the research program was a Harvard student spirometer. It consists of a counterpoised wedge-shaped float of capacity 9 l, hinged to operate within a water filled tank which effects a seal to the float. The tank and float are one piece mouldings of a rigid plastic material. A subject breathes through the system via a one-way air valve system. Large bore polyethylene tubing interconnects the mouthpiece and valve assembly to the apparatus via a two-position tap. Breathing patterns are recorded via the excursion movement of the float: A low inertia potentiometric transducer is coupled to the hinge of the float to allow monitoring of the float excursion. The potentiometer, connected as a potential divider, provides an output voltage change as a linear function of air volume change. As the transducer is

powered with a battery, the output signal maximum amplitude is only 300 mV. Such signal, after adequate conditioning[5], is sent to the FORP acquisition system for recording.

Because the lung residual volume cannot be measured with a volumetric method, the spirometer volume measurements are relative to the subject's residual volume. Calibrating the spirometer consists of 1) de-biasing the transducer output signal so that it indicates a specific known value at the residual volume and 2) finding the proportionality factor between the relative air volume and the signal amplitude.

An ideal spirometer would offer extremely little resistance to the movement of air to and from the subject's lungs so that it would not affect the subject's ventilatory movements. In practice, airflow resistance cannot be completely eliminated.

In its basic configuration, the spirometer is a closed system so the subject re-breathe his/her exhaled air. This contributes to the enrichment of carbon dioxide and impoverishment of oxygen in the closed finite air volume, resulting in the subject becoming short of breath and light headed with eventual loss of consciousness.

These effects are prevented by either 1) removing the carbon dioxide and adding oxygen or 2) shortening the duration of the acquisition or 3) measuring the volume of exhaled air only. Removing carbon dioxide can be accomplished by inserting a soda lime[6] canister in the airflow coming from the mouthpiece. In the same fashion, adding oxygen is made possible by inserting an intake of medical oxygen into the airflow going to the mouthpiece.

In order to reduce the airflow resistance, the spirometer must be kept to its simplest configuration. FORP experiments involving the acquisition of a spirometric signal were reduced to a duration of 30 seconds. This corresponds to the duration for which it is possible to deprive a healthy adult male subject from oxygen intake without exposing him/her to the risks mentioned above [Pocock et al (1999)].

The spirometer is both invasive and non-ambulatory, essentially restricting its zone of utilisation to the physiology laboratory. The instrument maintenance requires the regular cleaning of all tubing and eventually the sterilisation of the mouthpiece.

---

[5] Conditioning here signifies bias correction and amplification. Filtering is not required as the signal is almost noiseless.

[6] Soda lime is a carbon dioxide absorbent.

2.5.1.2. Pneumotachometer

The pneumotachometer permits the same lung volume measurements as with its direct counterpart the spirometer, the difference with the former is that pneumotachometers can be made ambulatory. Pneumotachometry is a mouth airflow measurement technique [Fleisch (1925), Sullivan et al. (1984)]. The pneumotachometer measures the flow of air exchanged between the lungs and the outside environment with a very low inertia turbine mounted on a face mask. The turbine's rotational speed (measured with an opto-electronic system) is related to the bi-directional airflow and a lung volume measurement is obtained by integrating the airflow in time. The pneumotachometer does not require extensive air tubing since the active part of the system (the turbine) is placed directly in front of the mouth, hence reducing the airflow resistance compared to a spirometer. However, a nose clip is required to prevent air circulation by the nose. The mechanical simplicity and reduced size of the system explains the possible adaptation of pneumotachometers as ambulatory systems.

2.5.2. Non-invasive monitoring

There are numerous ways to monitor respiration non-invasively. This is not the focus of this thesis to provide a detailed comparison of all methods but we will briefly present the most popular ones because we must know about the FORP concurrent methods.

2.5.2.1. Proximal airway pressure monitoring

Apart from torso movements, respiration is manifested outside the body by a bi-directional airflow from the mouth and the nose. Proximal airway pressure sensors maybe located in such regions and measure the change of pressure due to the airflow at the extremities of the airways. [Jensen (2001)]

2.5.2.2. Thermal respiratory monitoring

In most case, the airflow generated during exhalation has a different temperature than the airflow during inhalation (it's hotter). This difference can be measured with a thermistor[7], placed in front of the mouth or nose: this technique is known as thermal respiratory monitoring [Brown (1966)]

2.5.2.3. Acoustic respiratory monitoring

During ventilation, air bodies experience frictional forces on the inner walls of the airways and this may result in a local vibration of the ventilation system. Typically, snoring is the vibration of the glottis due to airflow during sleep. Acoustic respiratory monitoring is obtained by placement of a dedicated microphone on the neck of the subject. [Werthamer et al (1983), Henneberg et al (1992)]

---

[7] A thermistor is a dipole whose electrical resistance inversely related to absolute temperature.

2.5.2.4. Impedance pneumography

Impedance pneumography is the respiratory monitoring modality of a larger scope technique known as impedance plethysmography. As with all plethysmographic methods, impedance plethysmography is capable of cardiac monitoring. This is further detailed in section 8.2.

Impedance pneumography is based on the fact that air has a very low electrical conductivity compared to that of body tissues. Therefore, air volume variation in the lungs produces an electrical impedance variation of the chest. Impedance pneumography consists of measuring the variation of thorax impedance along the anteroposterior axis to estimate the respiratory volume variation. This is performed by circulating a 3 to 6 mA sinusoidal (100 kHz) current along the axis previously described and measuring the corresponding electric potential across the axis. The impedance signal is obtained from the complex ratio of voltage and current. [Allison (1964), Hamilton (1967)]

Impedance pneumography is very commonly used in clinical conditions because of its simplicity but one drawback of this technique is that the impedance change caused by blood flow, for instance by the aorta, can generate a signal of the same order of magnitude as impedance variations caused by respiratory activity. [Jensen (2001)]

2.5.2.5. Electrical Impedance Tomography

Electrical Impedance Tomography (EIT) is a non-invasive, cheap and fast, low spatial resolution imaging technique. It can be used to image respiratory movements if desired, but ultimately any part of the body can be imaged, including the heart.

EIT has evolved from impedance plethysmography: Its principle is to measure electrical impedance at the surface of the skin in multiple regions. Specific algorithm are being developed to construct a 2 D image of the section encircled by the sensor, this sensor is simply a belt equipped with multiple regularly spaced electrodes (about 12) for current injection and electrical potential measurements. The belt is connected to a computer dedicated to image construction.

2.5.2.6. Magnetometric respiratory monitoring

Magnetometric respiratory monitoring consists of estimating the respiratory volume by measuring the anteroposterior diameter of the thorax and abdomen with the help of a magnetometer. A magnet is placed on the dorsal wall of the thorax and a magnetometer monitors variations in the magnetic flux perceived at the skin surface of the anterior thoracic wall. Flux variations are a function of the thorax displacement due to respiration. [Konno and Mead (1967), Mead et al (1967)]

2.5.2.7. Photoplethysmographic respiratory monitoring

Photoplethysmography is a technique based on intensity measurement of light reflected by body tissues from an infrared optical source. The primary component of the signal is due to cardiovascular pressure waves but it was found that the signal also contains a second periodic component, which is of a respiratory nature [Lindberg et al. (1992)].

2.5.2.8. Plethysmographic respiratory monitoring

Since a majority of the studies within this thesis rely upon the plethysmographic method of determining breathing patterns, it is appropriate to comprehensively review its background and development. Plethysmographic respiratory monitoring relies on plethysmography to estimate lung volume by measurement of the subject's chest geometry. This technique is the standard for unobtrusive respiratory monitoring and has been used widely in clinical and research settings. Approximately 1600 published scientific studies have used this technology and established it as the standard for non-invasive assessment of breathing. The difficulties associated with this technique arise from the correct modelling of the integration of the ventilation system in the chest. This is thoroughly discussed in section 11.2.3. The Respiratory Inductive Plethysmograph (RIP) [Cohn et al (1975), Sackner et al (1980)] is the commercialised form of this concept and is the main counterpart of the FORP, which is presented in this thesis. Both systems are reviewed in detail in section 7.7.

Chapter 3 CARDIOVASCULAR SYSTEM AND CARDIAC MONITORING

3.1. Introduction

Respiratory plethysmographic instruments such as the RIP and FORP are capable of observing cardiac activity (cardiac movements and cardiovascular pulses) at different locations in the human body such as the thorax and abdomen [Jordan et al. (1984), Sackner et al. (1991), Babchenko et al. (1999a), Maletras et al. (2001b)]. This chapter intends to provide the necessary physiological knowledge of the cardiovascular system required to understand the origin of the cardiac activity observed by the FORP and the RIP under certain conditions.

At the end of this chapter, the electrocardiogram (ECG) system for cardiac monitoring is described and its use and results are discussed. The ECG machine has been employed in this research as a reference system to measure the time position of heartbeats.

Because the focus of this thesis is on the capability of the FORP to monitor respiratory and cardiac activities, plethysmographic cardiac monitoring systems, including the FORP, are presented in section 8.4.

3.2. Cardiovascular system

The blood circulatory system allows the cells of any animal body to exchange the product of their metabolism for oxygen and other nutriments via the circulating blood. As discovered in 1628 by William Harvey, the cardiovascular system forms a circle, so that blood pumped out of the heart through one set of vessels returns to the heart via a different set. There are actually two independent circuits, both originating and terminating in the heart, which is, for this reason, divided into two functional halves. Each half contains two chambers, the atrium (upper chamber) and the ventricle (lower chamber). In total, the heart is separated into four cavities: The right atrium (RA), the right ventricle (RV), the left atrium (LA) and the left ventricle (LV). The blood contained in the atrium on each side empties into the ventricle of the same side. There is no blood circulation between the two atria or the two ventricles. Blood is pumped via one circuit, the pulmonary circulation, from the right ventricle through the lungs and then to the left atrium. It is then pumped through the systemic circulation from the left ventricle through all the organs (except the lungs) and then to the right atrium. In both systems, arteries carry blood away from the heart and veins returns blood to the heart.

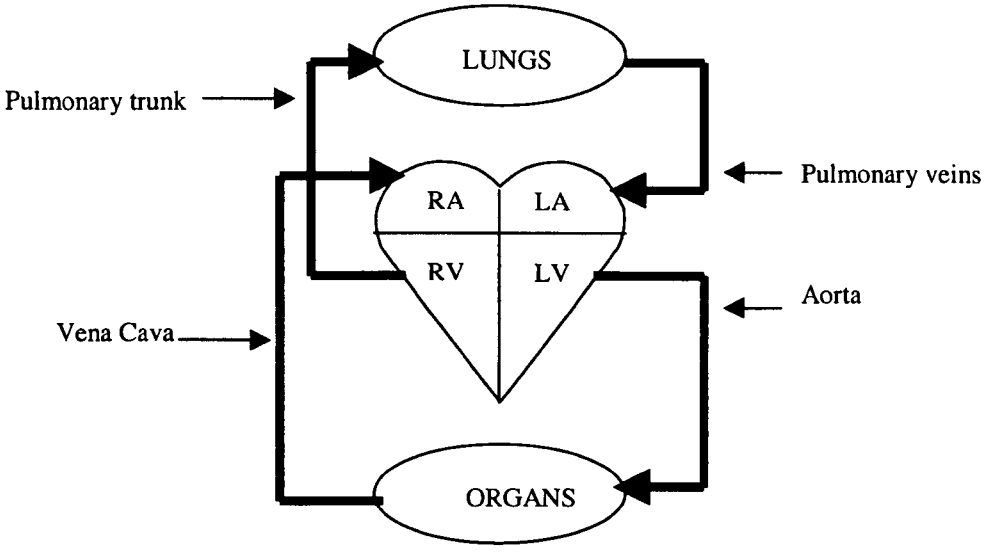

Figure 1:Simplified structure of the heart and cardiovascular system.

In the systemic circuit, blood leaves the left ventricle via a single large artery, the aorta. The arteries of the systemic circulation branch off the aorta, dividing into progressively smaller vessels. The arteries branch to the microcirculation formed of the arterioles, the capillaries and the venules, in decreasing order of magnitude respectively. The venules then unite to form larger vessels of increasing diameter called the veins. The inferior vena cava collects the de-oxygenated blood from the systemic circulation of the lower portion of the body and the superior vena cava assumes the same function for the upper half of the body. Both vena cava join in the right atrium.

In the pulmonary circulation, blood leaves the heart via a single artery called the pulmonary trunk, which divides into the two pulmonary arteries, one supplying the right lung and the other the left. In the lungs, the arteries continue to branch to smaller diameter vessels, ultimately forming capillaries, which are in contact with alveoli to process oxygen and carbon dioxide exchange between alveolar air and the blood. The capillaries then branch off to veins of increasing diameter. The blood leaves the lungs via four pulmonary veins, which connect to the left atrium. The blood in the pulmonary veins, left heart and systemic arteries has high oxygen content. As this blood flows through the capillaries of the systemic circulation, oxygen is extracted from the blood to the cells of the organs, resulting in the lower oxygen content of the systemic venous blood.

3.2.1. The heart anatomy

The heart of a normally constituted adult is a cone shaped organ about the size of his/her fist and weighs 250 to 300g. The heart lies close to the thoracic wall and is aligned on an axis defined by the left hip and the right shoulder.

Figure 2: Position of the heart in the ribcage [Fox (1999)].

The heart is a muscular organ enclosed in a fibrous sac called the pericardium, and located in the thorax. The space between the pericardium and the heart is filled with a fluid that serves as a lubricant as the heart moves within the pericardium. The walls of the heart are primary composed of muscles termed the myocardium. Located between the atrium and ventricle of each half of the heart is the atrioventricular (AV) valves, which permit blood to flow from atrium to ventricle but not reciprocally. The right AV valve is called the tricuspid valve and the left AV valve is called the mitral valve.

The opening and closing of the AV valves is a passive process resulting from the pressure differences across the valves. When the blood pressure in an atrium is greater than that in the ventricle separated from it by a valve, the valve is pushed open and flow proceeds from atrium to ventricle. In contrast, when a ventricle achieves a pressure higher than the atrium it is connected to, the AV valve between them is forced closed.

The opening of the right ventricle into the pulmonary trunk and the left ventricle into the aorta also contain valves termed the pulmonary and aortic valves. Like the AV valves, the pulmonary and aortic valves assume the same function and act in a purely passive manner by permitting blood flow out of the ventricles but preventing backflow. The closure of the AV valves and the pulmonary and aortic valves respectively is responsible for the double heart sounds.

There are no valves at the entrance of the atria, however atrial contraction pumps very little blood back into the veins because it compresses the veins at their sites of entry in the atria, greatly increasing the resistance to backflow.

The typical cardiac cycle is described in two phases, namely diastole and systole. During diastole, the AV valves are open and the ventricular volumes increase. At the end of diastole, the AV valves close and the systole begins with the opening of the pulmonary and aortic valves. The ventricular contraction leads to blood ejection and systole finishes with the closing of the pulmonary and aortic valves. A new diastole then starts with the opening of the AV valves.

During a cardiac cycle, ventricular contraction generates blood pressure variations in the cardiovascular system. In the systemic circulation system, the systolic and diastolic pressures are respectively 120 mmHg and 70 mmHg. In contrast, the pulmonary circulation system is a low pressure system where the systolic and diastolic pressures are respectively 24 mmHg and 8 mmHg. This difference in pressure between both circulatory systems is clearly reflected by the thickness of the left ventricle muscular structure compared to the right one. This difference in ejection pressure between left and right ventricles does not affect the volume flow because the cardiovascular system is a closed system: the volume flow in the aorta and in the pulmonary trunk are necessarily the same. Differences of pressure are smoothed out by virtue of the elasticity of the walls of the aorta and all other main arteries. This mechanism permits to maintain an uniform blood flow in the circulatory system at any time.

Important cardiological parameters are the cardiac rate, the stroke volume (SV) and the cardiac output (CO). The SV corresponds to the total volume of blood ejected from the heart during a single cardiac cycle and is calculated by the difference in heart volume between end diastolic period and end systolic period. The stroke volume is of particular interest in cardiology examination as it estimates the efficiency of the pumping action of the heart. Normal stroke volume for an adult at rest is about 70 mL per heartbeat. The cardiac output is defined as the total volume of blood ejected by the heart per minute and is calculated by the product of the stroke volume and the heart rate, assuming that the latest is constant.

3.2.2. Heart beat co-ordination

The heart is, in essence, an auto-rhythmical dual pump in that the atria contract first, followed almost immediately by the ventricles. The cardiac muscle cells of the myocardium are arranged in layers that are tightly bound together and completely encircle the atria and ventricles. Approximately 1 percent of the cardiac muscle cells does not function in contraction, but constitutes a network known as the conducting system. Such cells are able to receive and transmit electric potentials, in the manner of a simple electric wire, and deliver this potential to cardiac

muscle cells, via gap junctions in the cardiac muscles. The transmission of electric potential, or action potential, works by depolarisation of the cell membrane. Depolarisation is a change of equilibrium in the proportion of the potassium and sodium ions passing through the cells membrane. Repolarisation is the inverse phenomenon in which the membrane returns to its previous equilibrium position. At rest, the membrane potential is approximately –90 mV. Depolarisation increases the membrane potential to + 10 mV and repolarisation returns it to –90 mV. Consecutive depolarisation and repolarisation create a pulse shaped variation of the membrane potential of duration 250 ms. The rising side of the pulse (depolarisation) triggers depolarisation of the contiguous conducting cells and as a result, the potential variation is propagated along the conducting system. The sinoatrial (SA) node is constituted of cells whose membranes are periodically depolarised/repolarised in an autonomous fashion. In other words, it is the normal pacemaker of the heart and its discharge rate determines the cardiac rate: approximately 100 depolarisation per minute in the absence external nervous influence. A large number of sympathetic and parasympathetic nerves end in the SA node. Activity of the parasympathetic nerves causes the heart rate to decrease whereas sympathetic nerves cause the heart rate to increase. In the resting state, there is considerably more parasympathetic influence than sympathetic influence.

After the initial "firing" of the SA node, the action potential deploys onto the myocardium of the atria, progressing by way of gap junctions and therefore provoking both atrial chambers to contract. The atrial depolarisation eventually reaches the atrioventricular (AV) node, which acts as a relay and transmits the action potential to the rest of the conducting system. The atrial depolarisation cannot directly extend to the ventricles because a layer of non-conducting tissue separates atria and ventricles. The propagation of the action potential through the AV node is relatively slow (approximately 100 ms) and this delay allows atrial contraction to be completed before ventricular excitation occurs. After leaving the AV node, the action potential is guided along the conducting system by the AV bundle (also known as Bundle of His). The conducting system then separates into two branches, each guiding the action potential to respectively left and right ventricles. These branches in turn make contact with Purkinje fibres, large conducting cells that rapidly distribute the action potential throughout much of the ventricles. Finally, the Purkinje fibres make contact with ventricular myocardial cells, therefore causing depolarisation of right and left ventricles and consequently ensuring almost simultaneous ventricular contractions. In reality, the ventricular contractions begin at the apex of the heart and spread upward. The co-ordinated sequential contractions of atria and ventricles ensure maximisation of the ejected blood volume.

Figure 3:Innervations of the heart [Costanzo (2002)]. The black arrows represent the propagation of depolarisation.

Figure 4:Mean vector through the partially depolarised ventricles [Guyton et al (2001)]. The shaded area represents regions that are already depolarised (hence the negative signs all around). The diagonal vector indicates the axis of the largest maximum variation in electric potential during the cardiac cycle, which coincides with the axis of depolarisation. Potential variations across this axis can be recorded with the "lead II" electrode position of a three leads ECG system (see section 3.3).

3.3. Cardiac activity monitoring

Cardiac activity is among the most important vital signs, along with respiration. To monitor cardiac activity, we must deploy instruments that provide measurement of cardiac rate, cardiac pressure, and other cardiological parameters.

A range of non-invasive techniques, such as measurement of skin electrical potential (electrocardiography), measurement of skin optical transmission/reflection (pulse oximetry and photoplethysmography), measurement of body acoustic activity (stethoscopy), and measurement of body volume variation (plethysmography) can achieve cardiac monitoring. Aside from this, non-invasive imaging techniques such as CT, MRI or echocardiography are also widely employed during cardiological investigations.

A review of all existing techniques and instruments for cardiac monitoring is not the focus of this thesis. However, two of them are of particular interest to us: electrocardiography and plethysmography.

The electrocardiogram (ECG) technique is presented in the remainder of this chapter because it is the workhorse of the majority of cardiological investigations. In particular, we employed ECG recording in this research to characterise the presence of heartbeats during the development of cardiac monitoring with the FORP. The reason for choosing this technique is its reliability, in normal conditions.

Plethysmography is a technique equally capable of monitoring respiratory or cardiac activity. The volume variations of the beating heart produce thoracic tissues displacements and eventually thorax circumference variations [Vander et al (1994)]. Also, after each cardiac contraction, the heart emits a cardiovascular pressure pulse that propagates along the network of arteries and veins throughout the body. Such a pulse also generates local body volume changes. A fraction of the sum of these volume variations is eventually transmitted to the surface of the skin where a plethysmographic instrument can measure them.
Cardiac plethysmography is the main technique of interest in this thesis and, for this reason will be the object of chapter 8.

3.3.1. The electrocardiogram (ECG)

The heart contractions are triggered by propagation of electrical potentials through the heart walls by the conducting system into the myocardium of the atria and ventricles. The different electric currents generated by the heart during a cardiac cycle propagate through the whole body and eventually reach the surface of the skin. The skin has a relatively high electrical impedance and the potential differences due to cardiac activity observed at the skin surface are no more than a few millivolts in amplitude but can nonetheless by accurately recorded. The process of recording

cardiac electrical potentials is termed electrocardiography and results in a time signal called an electrocardiogram (ECG). The ECG signal perceived at the surface of the skin constitutes the weighted sum of all myocardial cells potentials. The typical shape of the ECG signal results from the sequence of events happening during one cardiac cycle, namely essentially atrial depolarisation (forming the P wave), ventricular depolarisation and atrial repolarisation (forming the QRS complex) and ventricular repolarisation (forming the late T wave). Since the origins of these events are well explained, the ECG is a powerful tool for the evaluation of heart condition. In this research, the ECG served as a reference technique for measuring the position in time of heartbeats, mainly so that their presence could be correlated with the variations in the FORP signals that were suspected to be of a cardiac nature.

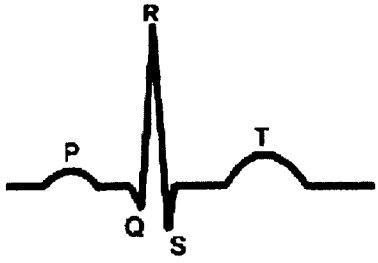

Figure 5: Schematic of a typical ECG signal. The P-R interval is taken from the start of the P wave to the start of the QRS complex. It is the time taken for depolarisation to pass from the SA node via the atria, AV node and His-Purkinje system to the ventricles. The QRS complex represents the time taken for depolarisation to pass through the His-Purkinje system and the ventricular muscles. The Q-T interval is taken from the start of the QRS complex to the end of the T wave. This represents the time taken to depolarise and repolarise the ventricles. The S-T segment is the period between the end of the QRS complex and the start of the T wave. All cells are normally depolarised during this phase.

3.3.1.1. ECG signal types

There are mainly two types of ECG recording namely 12 lead or 3 lead (bipolar ECG). A typical clinical ECG makes use of multiple combinations of 12 electrode locations on the limbs and chest so as to obtain as much information as possible concerning different areas of the heart. The shapes and sizes of the P wave, QRS complex and T wave varies with the electrode locations. The ECG can also be recorded with only 3 electrodes: in the bipolar configuration, electrodes are placed on the wrists and ankles and the different arrangement of electrode placements are called "Lead I", "Lead II" and "Lead III".

In Lead I, the signals originating from the left wrist (LA) and right wrist (RA) are fed to a differential amplifier[1] while the left ankle (LL) signal acts as an earth signal. In this arrangement, the ECG signal corresponds to electric potentials propagating between the left and the right halves of the heart.

In lead II, the differential amplifier is inserted between RA and LL and LA serves as earth signal. In this fashion, the ECG signal perceives the electric potentials travelling from the right atrium to the left ventricle.

Finally, in Lead III, the differential amplifier is positioned between the RA and LL signals, therefore producing an ECG signal representing potentials across the left atrium/ right ventricle axis. Lead II is the configuration we have used in this research. This is indeed the most commonly used [Lee (2000)] and ECG signals produced in this manner had good clarity and large magnitude.

Figure 6:Electrodes positions in the three leads ECG configuration [Fox (1999)]. Note that electric potential measurement across the lead II axis produces the largest variations. This is because the lead II axis is aligned with the direction of depolarisation (see also figure 4 in section 3.2.2).

[1] The differential amplification of electric potentials P1 and P2 would be $G*(P_2-P_1)$, with G being the amplification factor.

3.3.1.2. Limitations of the ECG system

Due to the electrical nature of the ECG signal, the ECG system cannot be used in a medium where harsh electromagnetic perturbations occur, such as the interior of an MR scanner, or in a medium likely to conduct electrical currents, such as rwater. Additionally, the ECG system presents a health and safety risk associated with the placement of electrodes directly on the patients skin: efficient electrical isolation is required to ensure that the patient does not experience electrical discharge from the electrodes through his/her heart.

This type of accident constitutes the primary source of concern in ECG monitoring during MRI: the strong EM fields in the scanner cavity may induce electrical currents in the ECG leads and eventually electrocute the patient to the point of severe burning at the location of the electrodes. Another common problem is the impedance change at the skin/electrode interface, provoking spontaneous variations in the ECG signal magnitude. In children, skin irritation due to ECG electrodes and paste are common complications. [Lindberg et al (1992)] The ECG system is also relatively sensitive to movements of the body during examination as the electrical activity associated with any muscular displacement is perceived in the ECG signal, therefore hindering the cardiac signal [Bazett (1920)].

More importantly perhaps, the ECG signal does not correspond to the actual cardiac movements but only reflects the cardiac activity as seen through the sequence of electric currents propagating from the heart to the surface of the skin. In other words, the analysis of an ECG signal only provides information which is used to estimate the muscular activity of the heart chambers.

In contrast, the FORP can measure cardiac volume variation qualitatively. This is further detailed in chapter 12.

Chapter 4 OPTICAL FIBRES

4.1. Introduction

This section aims at presenting the basic principle of the FORP sensor. This requires an understanding of optical power signal variations originating from the MacroBending Loss Effect (MBLE) in multimode step index optical fibres.

There are two approaches to describing optical systems: The ray optics approach and the wave optics approach. Ray optics is based on geometrical modelling of the propagation properties of light through different media whereas wave optics considers light as an electro-magnetic (EM) field whose existence and behaviour is dictated by Maxwell's equations. Ray optics has the inherent advantage of being simple to manipulate (or, at least, simpler than wave optics) and intuitive to understand. Wave optics, despite its complexity, provides a larger scope of understanding of optical phenomenon and encompasses ray optics.

Optical fibres, like any other optical system, can be explained using either the ray or wave approach. In the case of the FORP, we will always use fibre core diameters that are considerably larger than the light wavelength, and for that reason it is acceptable to describe light propagation in an optical fibre using ray optics [Senior (1984)].

Optical fibres consist of a rod of optically transparent dielectric medium, usually glass or plastic, where light is transported with minimal loss of energy over long distances by the physical process of Total Internal Reflection (TIR).

4.1.1. Step index optical fibre

The first main class of optical fibre is the so-called step index optical fibre. The principle of light channelling with TIR in step index optical fibres can be easily explained with ray optics. In the step index fibre, the fibre's internal rod, called the core (with refractive index n_1), is immediately surrounded by a cylindrical layer called the cladding (with refractive index n_2) which is lower than n_1. Light propagates in the fibre core by successive reflections at the core/cladding interface. Because $n_1 > n_2$, Snell's law states that no optical power will be transmitted to the cladding during the reflections if incident light rays possess a propagation angle θ greater than the critical angle θ_c.

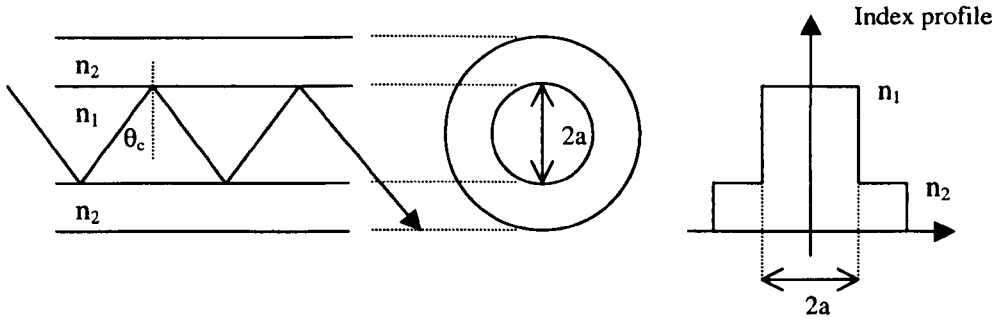

Figure 1: Two-dimensional approximation of the mechanism of light propagation in a step index optical fibre. The right diagram represents the refractive index variation as a function of the fibre radius.

The propagation angle θ exists in the plane defined by the incident light ray and the normal to the core/cladding interface. The condition for TIR, derived from Snell's law is expressed below:

$$\theta_c = \sin^{-1}(\frac{n_2}{n_1}) \qquad \text{(Eq. 1)}$$

This relation also expresses the underlying fact that there is an angular condition on the light entering the core at the fibre end: If light propagation by TIR along the fibre is limited by an angular condition then not all light rays entering the fibre core can be propagated. Light entrance is categorised by an acceptance cone, i.e. a truncated cone whose top is aligned with the core axis. This cone has a solid angle θ_{max} called the maximum acceptance angle. θ_{max} depends on the critical angle θ_c, hence on the refractive indexes of the core and cladding. θ_{max} is defined as:

$$\theta_{max} = \sin^{-1}(\frac{1}{n}\sqrt{n_1^2 - n_2^2}) \qquad \text{(Eq. 2)}$$

where n is the refractive index of the external medium at the fibre end. This is usually air, and is usually approximated as 1 in that case. However, to eliminate the dependence of θ_{max} on n, it is most of the time defined in terms of the Numerical Aperture (NA) of the fibre [Hecht (1998)]. The NA is defined as:

$$NA = n \cdot \sin(\theta_{max}) = \sqrt{n_1^2 - n_2^2} \qquad \text{(Eq. 3)}$$

NA is usually in the range 0.2 to 0.5 for most glass or optical fibers.

Light rays travel down the fiber as modes. Each mode is a group of rays characterised by a discrete propagation angle θ. [Ghatak and Thyagarajan (1998)] Modes have propagation angles in the range $[0:\theta_{max}]$ but the number of possible modes is a finite value. If a fibre can only accept 1 mode, it is termed monomode, otherwise it is termed multimode. The number of guided modes is dependent

upon the physical parameters of the fiber and the wavelength of the transmitted light. (The term guided modes refer to light paths whose energy is predominately transmitted within the fibre core, as opposed to unguided modes and leaky modes whose energy is spread in the cladding. These are, however, concepts arising from a wave optics treatment). The approximate number of guided modes in a step index multimode fibre [Senior (1984)] is given by:

$$M_S = \frac{V^2}{2} \qquad \text{(Eq. 4)}$$

where V is the dimensionless waveguide parameter, also sometimes called the normalised frequency. The expression for V [Ghatak and Thyagarajan (1998)] is given below:

$$V = \frac{2\pi}{\lambda} a \cdot NA \qquad \text{(Eq. 5)}$$

where λ is the light wavelength, a is the fibre core radius and NA is the numerical aperture. The parameter V allows us to calculate the number of modes that a fibre can carry at the same time. The expression above, is derived from wave optics but can also be understood in terms of ray optics.

4.1.2. Graded index optical fibre

The second main class of optical fibres is termed graded index optical fibre. The refractive index varies smoothly from the centre of the core to the core-cladding interface, in a radially symmetric manner. Usually this variation in refractive index is approximately parabolic. In this case, the light path is no more a saw-teeth pattern (see figure 1), but a sinusoidal curve (see figure 2). The amplitude of the sinusoidal path depends on the initial incidence angle of the light ray: the larger the angle, the bigger the amplitude and, subsequently, the longitudinal distance covered by all modes is almost identical. In other words, all modes cover the same distance at the same speed and therefore, the propagation delay of the graded index fibre is almost constant for all modes, occasioning only limited intermodal dispersion (see section 4.2.1).

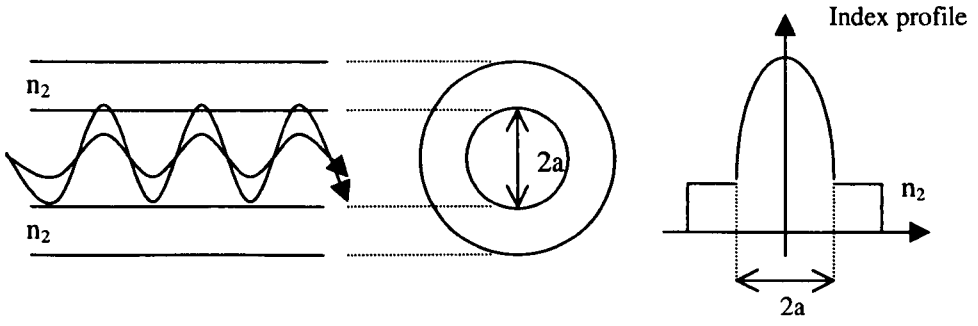

Figure 2 Two-dimensional approximation of the mechanism of light propagation in a graded index optical fibre. The right diagram represents the refractive index variation as a function of the fibre radius. Note the parabolic shape.

Graded index optical fibres have not been used for the construction of the FORP sensor for the reason that they do not produce sufficient signal attenuation under the effect of macrobending loss [Ghatak and Thyagarajan (1998)]. The Macro-Bending Loss Effect (MBLE) is discussed in section 4.3.4.

4.2. Dispersion

In the FORP system, optical fibers are energised with a pulsed light source. The source is pulsed because one can temporarily obtain more power from the light source during a short time interval than in continuous mode. At any time during the duration of the pulse, the total optical power is the sum of the power in all modes. At the launch point into the fibre, the pulse can be represented by a simple rectangular signal in a power versus time graph.

Upon transmission of such a light pulse through an optical fibre, the pulse received at the end of the fibre can appear smeared. This phenomenon can be explained in terms of intermodal and spectral dispersion. We are interested in quantifying the dispersion of the pulses transmitted through the FORP sensor as such dispersion could potentially affect the efficiency of the optical power measurement.

4.2.1. Intermodal dispersion

Because all modes have different propagation angles, they all describe a different geometrical path. For the multimode step index optical fibre, it is easy to visualise that each path has a specific length, or spatial period. Different length with constant speed implies different propagation time. The higher the propagation mode, the longer the length, the slower the mode. Consequently, when different modes are launched simultaneously into the fibre, they will reach the other end of the fibre at a different time and there is a dispersion in the time of arrival of each mode. This is the phenomenon of intermodal dispersion and it is inherent to multimode step index optical fibres. A consequence is that by modifying the distribution of power in time, intermodal dispersion will spread a Dirac shaped pulse of light at the fiber entrance into a Gaussian shaped pulse at the fiber

output. The intermodal dispersion time is the time difference between the arrival at the fibre output of the fastest and the slowest ray. It is given by:

$$\Delta t = \frac{Ln_1}{c}\left(\frac{n_1}{n_2} - 1\right) \qquad \text{(Eq. 6)}$$

where Δt is the dispersion time, L is the optical fibre length, c is the speed of light, n_1 is the core refractive index and n_2 is the cladding refractive index. In the case of the FORP, L is less than 10 m and it follows that the intermodal dispersion delay is less than 1 ns i.e. less than 0.01% of the pulse duration (10 μs). Such small broadening no influence on the signal detection because the signal loss produced by the pulse spreading is negligible.

The graded index optical fibres have the interesting property that they have very limited intermodal dispersion, because the refractive index gradient produces a speed gradient that retards low order modes relative to high order modes. All modes remain effectively in phase along the fibre.

4.2.2. Spectral dispersion

Another source of discrepancy in mode travelling time is the spectral (or chromatic) dispersion. In the visible region of the EM field, the refractive index n of most normal media usually decreases with the wavelength. Also, we know that the speed of light propagation v into a medium of refractive index n is given by:

$$v(\lambda) = \frac{c}{n(\lambda)} \qquad \text{(Eq. 7)}$$

where c is the speed of light in vacuum and λ is the light wavelength. Because n increases slightly with increasing λ in the visible region, it follows that v decreases with λ, hence creating dispersion in the time of travel of different wavelengths covering the same distance.

This dispersion in time is termed spectral dispersion. Naturally, it can be limited by ensuring that the optical source illuminating the optical fibre aperture is as monochromatic as possible.

To quantify the spectral dispersion, we use the following example [Ghatak and Thyagarajan (1998)] of a silica fiber illuminated with an 850 nm wavelength LED of bandwidth 30 nm, which results in a spectral dispersion of 2.7 ns/km. For comparison, a similar LED of wavelength 950 nm and bandwidth 55 nm illuminates the FORP sensor optical fiber. Given that the fiber has a length of less than 10 meters and that the pulse duration is 10 μs, we conclude that it is safe to ignore the effect of spectral dispersion.

4.3. Transmission loss mechanisms

The proper channelling of a light signal through an optical fibre is heavily dependent on the fibre's ability to establish and to maintain the conditions of TIR. Any inability to sustain TIR results in signal attenuation.

Optical signal attenuation within optical fibres, as with metallic conductors, is usually expressed in the logarithmic unit of the decibel (dB). The attenuation is used to compare two power levels and may be defined for a particular optical wavelength as the ratio of the input (transmitted) optical power P_1 into a fibre to the output (received) optical power P_2 from the fibre as:

$$A = 10 \cdot \log\left(\frac{P_1}{P_2}\right) \qquad \text{(Eq. 8)}$$

where A is the attenuation in dB. P_1 and P_2 can be measured with an optical power meter.

The inability to sustain TIR relates to irregularities along the optical fibre, such local variations in refractive index, alignment of the fibre, variations in the core radius, and irregularities at the core/cladding interface. [Senior (1984)]

When a light ray encounters such an irregularity, the propagation angle of the light ray is modified and a different propagation mode is adopted. This is the phenomenon of mode coupling. In multimode fibers, not just one ray but a group of rays is coupled. This results in a change in the Mode Power Distribution (MPD) of the fibre. The MPD is a function of the mode power density versus mode propagation angles. [Boechat et al. (1991)]

As previously stated, the propagation angles occupy the range 0 to θ_c. If mode coupling results in shifting a propagation angle θ above θ_c then such a mode is lost by eviction from the core into the cladding.

Experimentally, we observe that the main cause of signal loss is due to termination of the fibre as they are very sensitive to misalignments. Other causes for signal loss are material absorption, linear or non-linear scattering and finally bending loss.

4.3.1. Material absorption

Material absorption loss results in dissipation of optical power as heat due to the material composition of the fibre core. The dissipation may be intrinsic (caused by the interaction of light with one the components of the fibre) or extrinsic (caused by the interaction of light with impurities of the fibre materials.)

4.3.2. Linear scattering

Linear scattering losses are due to physical imperfections of the fibre and are related to the quality of the fibre manufacturing process. Scattering refers to optical power being shifted from one propagation mode to another, often with the consequence that the other mode can no longer be propagated in the fibre core and is radiated away. Rayleigh scattering is a linear scattering process generated by embedded imperfections originating from temperature inhomogeneities during the cooling period of fibre fabrication. Mie scattering is another linear scattering process where losses are due to physical inhomogeneities of the order of a wavelength at the core/clading interface.

4.3.3. Non linear scattering

Non-linear scattering losses relate to the non-linearity of optical power transmission for large optical power input. The two main mechanisms of non-linear scattering are stimulated Brillouin scattering and stimulated Raman scattering. Non-linear scattering loss is more likely to happen in monomode fibre. The FORP sensor is immune to this loss effect because it does not rely on monomode fibre but multimode fibre, but more importantly, because the optical powers used are well below the threshold required to produce non-linear effects. [Senior (1984)]

4.3.4. Macrobending loss effect

The macrobending loss effect (MBLE) is a transmission loss phenomenon underlying the FORP sensor principle. It results in mode coupling through curvature of the optical fibre. In the FORP sensor, we provoke MBLE deliberately. A knowledge of the MBLE helps in controlling and maximising it, and a description of this phenomenon follows.

So far, we have assumed that the fibre was straight but optical fibres can be bent to some extent because the materials they are made of are not completely rigid. This is true for plastic fibres but it also applies to silica fibres. When the fibre core is bent (curvature radius in the range 1 to 10 cm), mode coupling occurs and the propagation angles of all modes are modified: the mode population is disturbed. Depending on the amount of bending, a particular mode is either coupled into another mode or evicted completely from the fibre core. Macrobending loss analysis is concerned with studying the modification of the MPD due to substantial bending of the optical fibre.

In the research work concerning the new FORP sensor presented in section 9.4, we are interested in measuring how much optical power[1] is lost from a step index multimode fibre by bending it.

Graded index optical fibres, whose core refractive index is a parabolic function of the core radius, produce light paths that are gradually curved because they are continuously refracted. Graded index fibers show a smaller dependence on macrobending [Ghatak and Thyagarajan (1998)] and for this reason they are not reviewed in this section because they have not been used in the development of the FORP sensor.

The MBLE has received considerable attention because of its adverse effects, especially in fibre optic telecommunication and in large power delivery system. In the typical geometrical representation of a step index core optical fibre, light rays travel at particular angles, each identified as a propagation mode. Illuminating the optical fibre entrance with an incoherent source (typically, an LED) results in producing propagation modes of high order that are totally internally reflected not only at the core/cladding interface but also at the cladding/jacket interface. Such modes are not strictly speaking guided modes as they travel through the core/cladding interface and back. These modes constitute the so-called evanescent wave.

The MBLE modulates the mode power distribution. During MBLE, propagation modes undergo two phenomena: coupling and/or eviction. This results in the mode population becoming reordered. As the fibre curvature radius decreases, modes that are most likely to be eradicated first are those in the evanescent wave, as they cannot be coupled back into the fibre once they have been transmitted through the core/cladding interface.
A wave optic explanation of the MBLE is given as follows: As the wavefront can only be planar, even inside a bend, the local wavefront velocity must increase at the outside of the bend to compensate for the larger distance to be covered. When the mode velocity reaches the speed of light in that medium, modes can no longer be guided and their energy is radiated away in the cladding.

Various contributors have investigated the topic of MBLE in the last thirty years. Different MBLE models exist that were derived either from ray optics [Jones (1987)] or wave optics [Thyagarajan et al. (1987)] but a generally accepted model has not emerged so far: the calculation of MBLE is a complex topic and different existing models yield different results.

The simplest approach, however, is to model the measurement of the optical power attenuation versus bending radius. Ramsay and Hockam (1980) have proposed the use of the following phenomenological equation as a MBLE attenuation coefficient:

---

[1] The optical power is the summation of the MPD over all propagation angles. The optical power is received and measured by a photodetector at the fibre output.

$$\alpha_T = A \cdot e^{-B \cdot R} \qquad \text{(Eq. 9)}$$

where α_T is the attenuation coefficient defined as the ratio of output optical power over input optical power, A and B are calibration constants and R is the bend radius of the fibre. This phenomenological equation has the advantage of being extremely versatile and therefore adaptable to any MBLE measurement but, because the equation has to be calibrated (using least squares for example) for each measurement, it cannot be used to forecast MBLE analytically. Essentially, a phenomenological equation can give results in good agreement with actual measurements but does not offer physical reasons for the behaviour of these measurements.

A more physical approach is provided by Boechat et al., based on previous work by D. Gloge (1972), who derived a wave optics model of the MBLE. This model has advantage of yelding simple MBLE calculations, despite its initial complexity. It is now presented in details.

In Gloge's initial model, the cladding is assumed to be of infinite size and only high order modes are considered. The model is based on a coefficient of power loss attenuation per unit length of circular path. This coefficient is a function of the mode propagation angle and the fibre curvature radius. It reflects the attenuation of the electric field in the cladding during bending loss. As can be seen in the equation below, the coefficient has an exponential expression that is a strong function of the mode propagation angle θ.

$$\alpha = 2nk\left(\theta_c^2 - \theta^2\right) \cdot e^{-\frac{2}{3}nkR\left(\theta_c^2 - \theta^2 - \frac{2a}{R}\right)^{\frac{3}{2}}} \qquad \text{(Eq. 10)}$$

where k is the angular wave number $2\pi/\lambda$, n is the core index, θ_c is the critical angle, a is the core radius and R is the fibre radius of curvature. By inspection, it is easily shown that all modes above an apparent critical angle θ_f are quickly eliminated. An expression of the apparent critical angle θ_f is given by:

$$\theta_f = \theta_c \sqrt{1 - \frac{2a}{R\theta_c^2}} \qquad \text{(Eq. 11)}$$

The expression above is very important as it expresses the modification of the angular condition for TIR at the core-cladding interface due to fibre curvature. In essence, the MBLE generates an apparent increase of the real critical angle due to the bend. Below is given the example of a guided mode propagating with angle θ_c in a straight fibre, which suddenly becomes extinguished when the condition for TIR cannot be satisfied at the entrance of the bend.

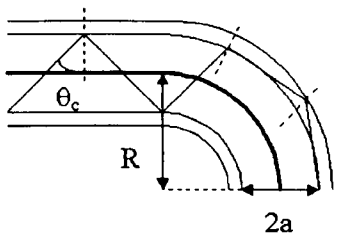

Figure 3: Demonstration of the increase in critical angle due to curvature

Boechat and co-workers (1991) have derived an approximation of the relative power attenuation by MBLE based on the Gloge model. The approximation is adapted for short length (circa 5 m) of large core (>0.2 mm) multimode fibre, which corresponds to the conditions of the FORP sensor. The Boechat approximation uses the mode power distribution of the fibre at the optical launch point. For such a short fibre, it is assumed that natural mode coupling (i.e. coupling that occurs without fibre bending) does not occur sufficiently for the mode distribution function to reach equilibrium. The distribution is therefore dependent on experimental conditions such as the launching optics. The distribution profile cannot be derived theoretically, it has to be measured and eventually modelled for simplicity of use. After measurement, modelling work by Boechat indicates that the best approximation of the profile for large core multimode fibre can be approximated by a trapezium.

To calculate the relative power attenuation, the coefficient α is integrated with respect to the circular optical fibre path length and multiplied by the MPD. The resulting expression is then summed over all modes to provide a global power attenuation coefficient. The relative power attenuation P_r is defined as the difference in global mode power before and after a curvature divided by the global mode power before the curvature. An expression for P_r is given by:

$$P_r = \frac{\int_0^{\theta_c} P_0(\theta)d\theta - \int_0^{\theta_f} P_0(\theta)\exp(-\alpha d)d\theta}{\int_0^{\theta_c} P_0(\theta)d\theta} \qquad \text{(Eq. 12)}$$

where θ is the mode propagation angle, θ_c is the critical angle, θ_f is the apparent critical angle, $P_0(\theta)$ is the MPD at launch point and α is the coefficient of power attenuation per unit length of circular path. Since α varies extremely rapidly with θ and most modes beyond θ_f are completely extinguished, the function α is better described as a switch (and therefore can be simplified as a step function) than a modulating factor. Using this characteristic, Boechat proposes a simplified relative power attenuation coefficient P_r that does not contain the attenuation coefficient α. An expression of the approximation for P_r is given in equation 13:

$$P_r \approx \frac{\int_{\theta_f}^{\theta_c} P_0(\theta)d\theta}{\int_{0}^{\theta_c} P_0(\theta)d\theta} \qquad \text{(Eq. 13)}$$

where θ is the mode propagation angle, θ_c is the critical angle, θ_f is the apparent critical angle, $P_0(\theta)$ is the MDP at launch point. Despite the absence of α the simplified relative power attenuation produces minimal differences (between 0 to 0.5%) to the original power attenuation expression. This justifies the approximation. Boechat also indicates that the curvature length has a negligible effect on the bending loss compared to the radius of curvature itself: the approximation does not include any reference to the fibre curvature length l. This suggests that most of the MBLE takes place at the entrance of a circular bend.

The behaviour of P_r is calculated from the MDP and the apparent critical angle θ_f. The expression for θ_f was originally associated with α and therefore θ_f is the only remaining contribution to the original model by Gloge in the approximation of P_r. By inspection, we can say that θ_f decreases extremely rapidly for small curvature radius, therefore quickly expanding the integration domain towards small propagation angles when R diminishes. This sudden rapid expansion of θ_f creates a similarly rapid and sudden increase in P_r for small curvature radii. This emphasises the importance of the accuracy of the MDP when calculating P_r for small curvature radius.

In the case of the Boechat experiment, measurement of the MPD of a laser source reveals that it can be approximated by a narrow top trapezium, in other words that a large proportion of all the mode power population is constituted of near axial modes. P_r is calculated with this trapezium approximation and then compared to actual measurements. The comparison reveals that the calculation is in good agreement with the measurement but that the model overestimates the MBLE for small curvature radii. One must remember that the Gloge model was developed with the idea of calculating high order bending loss and therefore is extrapolated to low order modes without guarantee of accuracy.

The MPD in the first few meters of the fibre is uniquely dependent on experimental conditions and essentially on the optical components being used. For high quality optical fibre systems with good metallic optical connectors, one can assume that the measurement of the MPD will be fairly consistent. This is not necessarily the case for lower quality plastic connectors such as the ones that were used in this research work (Siemens SFH 450 V and SFH 250 V). Such connectors, because of their high mechanical tolerance, do not provide consistency in the fibre connections and, as a consequence, alter the MPD by "leaking light".

For this reason, the Boechat approximation is strongly system dependent and therefore not transportable. In other words, it can be used to assess the MBLE of an existing optical fibre but not to forecast the loss of a system to be built, prior to MPD measurements.

Whatever the method being used, the MBLE calculation problem will remain strongly dependent on experimental parameters and no purely theoretical model can deliver a quantitative answer to how much power is lost during curvature.

The output of the MBLE model has been compared against real MBLE measurement obtained from the new FORP sensor developed in this research program. This comparison of experimental and theoretical responses is reported in section 9.5. It clearly shows that model and experimental results are in agreement down to a fibre curvature radius of 2 cm, the relative error at this stage is about 1%. As the curvature radius decreases, the relative error between the two responses increases dramatically: 15% at 1.5 cm and more than 30% at 1 cm.

Chapter 5 DIGITAL SIGNAL PROCESSING

The term Digital Signal Processing (DSP) is used to describe the complete set of operations, arithmetic calculations and numerical manipulations which are performed by a digital computer on the group of numbers representing the signals to be processed, in order to produce another set of numbers representing the processed signals. Many different functions can be performed in this way: spectral analysis, linear or non-linear filtering, encoding, decoding, modulation, detection, extraction, estimation of statistical parameters, etc.

One aspect of the research work presented in this thesis is the interest in developing algorithms and filters to break down FORP signals into sub-signals (or signal components) of different frequency ranges. The development of DSP algorithms for the treatment of FORP cardiac and respiratory signals involved the knowledge of some statistical concepts in signal classification, mathematical tools for signal analysis in both time and frequency domains and the principles of digital filters. These elements are given in the present chapter.

5.1. Need for DSP

When observing a physical phenomenon with a measurement instrument, we sometimes realise that the signal (commonly named "observation") contains additional information that might not be desired. Consequently, we start thinking about ways to extract only the desired information from an observation. The non-required information is referred to as noise, whether it actually is real noise or not.

To remove noise from a signal, one usually converts the observation into an electrical signal that can be processed by a denoising analogue circuit. Such a circuit would traditionally condition and bandpass the noisy signal to clean it up. This approach is a milestone in signal processing and presents many advantages, one of them being that analogue processing of signals naturally operates in real time.

However, the number of different operations that can be deployed with an analogue signal processing circuit is limited compared to what can be performed with a digital computer. Digital computers can carry out a virtually infinite number of computational tasks involved in signal processing, the only real limit being the computation time allowed for each task.

Digital computers are complex physical systems that realise logical and mathematical operations. They can be programmed to simulate systems that behave in a completely non-physical manner. For example, in the analogue domain, filtering operations are necessarily recursive[1] procedures. Analogue filters can be exactly reproduced in the digital domain by writing programming code that mimics the analogue filter's behaviour.

---

1 A recursive filter is a filter whose present output value not only depends on the past and present input values but also on the past output values.

Because the analogue filter is only simulated, if we should decide that there is no need for realism, we could then freely decide to modify the behaviour of this simulation: typically, one can think of a filter that does not use recursion. Such a non-recursive filter (also a called Finite Impulse Response filter) presents extremely interesting characteristics, such as a linear phase variation to quote only one, but would have no equivalent in the analogue domain as it ceased to represent a physical system.

Another major advantage of digital computation is the capacity for a digital system to permanently store the computation results in the computer memory, offering perfect reproducibility of the results.
To benefit from the advantages of digital computation in signal processing, analogue signals must be converted into a format that can be understood by the digital computer.
Conversely, one might want to pass digital data into the analogue domain. These tasks require conversions from the analogue domain into the digital domain and vice versa.

5.2. Conversions between analogue and digital domains

The analogue domain is concerned with continuous signals generated by physical phenomena and the digital domain is the recipient of stand-alone numerical values that represent the analogue signals. Here we review the operations involved in analogue-to-digital (ADC) and digital-to-analogue (DAC) conversions by taking the example of a one dimensional continuous analogue signal s with parameter t for time. Such analog signal is designated as s(t).

5.2.1. Analogue to digital conversion

The conversion of an analogue signal to a numerical form involves a double approximation. First, in the time space, the continuous signal s(t) is periodically sampled every time interval T, the sampling period.

The signal s(t) is then reduced to a set of real values equally spaced in time designated by s(nT), n being the integer time index addressing a particular value of the signal. Following sampling, each value of s(nT) is approximated (or quantified) by a multiple of an elementary quantity called the step. The value obtained by quantification is associated with the index of the concerned step in a process called coding.

The height of one step is the smallest amplitude range that can be resolved by the quantification process. The total number of steps is a power of 2 of the ADC resolution. In the FORP system, we have used the 12 bit industry standard resolution, giving a total number of steps of $2^{12} = 4096$.

Shannon's sampling theorem states that for a signal to be successfully sampled, the sampling period must be at least half the period of the minimum period signal component. This implies that sampling cannot properly resolve signal components with a period smaller than twice the sampling period. It follows that digitised signals have their frequency range limited to half the sampling frequency. It can be shown that the spectral resolution, i.e. the smallest frequency interval that can be resolved in a digitised signal is the inverse of the signal length.

5.2.2. Digital to analogue conversion

The process of digital to analogue conversion is performed when digital computation must be transferred back to the analogue domain. This operation is realised by decoding the step number to assign it a voltage value and interpolating between two consecutive voltage values to provide continuity in-between.

5.3. Typical digital signal processing system

The typical DSP based measurement instrument has a chain structure reflecting the flow of information carried by analog and digital signals. This structure is as follow.

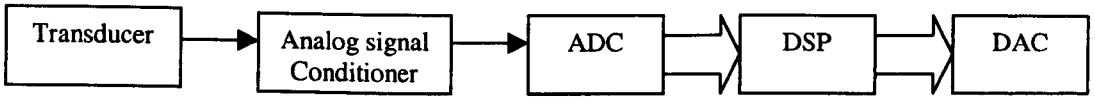

Figure 1:The typical DSP based measurement instrument. Narrow arrows represent analog signals and large arrows represent digital signals.

5.4. Statistical classification of signals

Signals can be divided into two main groups: deterministic and random signals [Bendat and Piersol (1986)]. Signals representing deterministic phenomena are exactly described mathematically and can be exactly categorised as periodic or non-periodic. Periodic signals are further divided into sinusoidal signals and complex periodic signals, such as the harmonics of the Fourier series. All possible ratios of the frequencies of such harmonics must form rational numbers. If this is not the case then the signals are not periodic but almost periodic. Almost periodic signals belong to the non-periodic group. Other members of the non-periodic group are transient signals.

Random signals, on the other hand, are not explicitly described by mathematical equations and correspond to measurement data representing a physical phenomenon. Random signals are never identical because each observation (i.e. each signal) of the phenomenon is unique. However, they can be mathematically modelled in many cases. A random process (also called a stochastic process) is an ensemble of real valued functions, which can be characterised through its probability structure. By probability structure, we understand the different moments (see section 5.5.5) of the signal and its probability density and distribution functions (see section 5.5.1) [Bendat and Piersol

(1986), Lynn (1982)]. Random signals are further divided into two groups: Stationary and non-stationary signals. A so called stationary random process is observed when the mean value (i.e. the hypothetical value) of a new signal at time t_1 can be accurately estimated by averaging the values at t_1 of multiple signals previously acquired and originating from the same phenomenon. In other words, stationarity expresses the idea that all signals generated at different times by a particular phenomenon share a common behaviour or a resemblance. If random signals do not directly have a common history (for example because an external phenomenon might interfere with the phenomenon under study) then the signals are labelled non-stationary. In practice, non-stationarity is difficult to investigate, unless one can make assumptions about the source of non-stationarity.

Random stationarity can further be classified in terms of ergodicity: If a time region of a random stationary signal shares the same statistical properties as the entire signal, then the signal is said to be ergodic. In other words, ergodic signals have statistical properties that remain constant, regardless of the size of the time region under observation. Consequently, the statistical properties of only one time region is required to represent the statistical properties of the entire signal.
This is very important because it sets us free from having to use multiple observations to extract relevant information from a process: only one observation is required and consequently, the required information can potentially be extracted in real time [Reynaud (1995)].

5.4.1. Statistical classification of FORP signals

In the research work that we present here, the statistical classification type of FORP respiratory signals is debatable because the signals are neither clearly stationary nor clearly non-stationary. This is due to the fact that respiration stops being an autonomous periodic reaction when active mind control is taken.
For example, the respiratory signal is non-stationary during talking and mostly stationary during sleep. Fortunately, the stationarity problem can be referred to the duration of the observation: A 30 second long respiratory signal is more likely to stationary than a 60 minute one because body activity and breathing style is very likely to remain constant over short periods. In the 30 seconds duration of the acquisition period, and under the conditions of the respiratory signal acquisition experiment described in chapter 11, we have observed that the respiratory signals are stationary. See section 11.2.2, figures 2 to 5.

However, their statistical properties differ from one another and consequently they are classified as non-ergodic. This makes it more difficult to algorithmically process these signals (for respiratory rate measurement for example) because of the irregularities they contain. These irregularities might be understood in terms of low-frequency, narrow-band noise. Cardiac signals however, observed with the FORP during apnea or ventilation are clearly stationary and ergodic in nature, regardless of the duration of the acquisition session. See section 12.4.1.4, figures 9 to 14.

5.5. Statistical methods for analysis of discreet signals

In order to quantify the statistical properties of a discreet signal, we use a set of statistical tools (derived from the analysis of continuous signals) to reduce the studied signal into meaningful quantities.

The most widely used tools of statistical analysis are the mean , the variance, the autocorrelation function and the power spectral density.

Let the variables x and y represent two simultaneously acquired discrete real-valued random signals of N samples each. In the following, the line underneath the method's equations is the equivalent formulation in Matlab script.

5.5.1. Probability Density Function

The Probability Density Function (PDF) of a signal permits to measure of the probability to find a particular group of values in a signal. When using discreet signal, the PDF can be approximated by the histogram, if the number of histogram bins is made equal to the range of signal values. Dividing the histogram by the number of values yields the approximated PDF. In Matlab script, this is expressed as:

$$x\_pdf = hist(x,nb)/N; \quad \text{(Eq. 1)}$$

where x_pdf is the approximated PDF, nb is the number of bins in the histogram, and N is the number of samples in the signal x. For x_pdf to be the closest possible approximation of the PDF, the number of bins nb must be made equal to:

$$nb = round(max(x) - min(x)); \quad \text{(Eq. 2)}$$

The PDF is often used to statistically define a separation (or threshold) between two amplitude regions of the signal. For example, we might want to know where is the most "popular" amplitude region. To measure this "popularity", or probability of occurrence, we would need to 1) obtain the PDF of the signal, 2) inspect the PDF to locate the maximal probability and, 3) determine the range of values associated with this maximal probability. Our threshold would be on the edge of this range.

Since the probability to find a signal between its minimum and its maximum is certain, the representation of the PDF has a unit area.

It is worth noting that the PDF of a signal tells us nothing about its detailed structure, nor is it a unique property of a signal: Different signals can share the same statistical properties, and this is also valid for the PDF.

5.5.2. Mean

The mean (or average, or expectation) is used to quantify the central position of a whole data set. There are two ways of calculating it. The first is by taking the sum of the product of the PDF and the span of all possible values of a signal. This is only possible if the PDF is known. The second and more practical way of calculating the mean is by taking the sum of all samples in the discreet signal and dividing the result by the number of samples. This is given by:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = E(x) = \mu_x \qquad \text{(Eq. 3)}$$

x_mean = mean(x);

5.5.3. Deviation

The deviation is what remains of a signal when its mean has been subtracted. The signal's deviation is given by:

$$d(x_i) = x_i - \bar{x} \qquad \text{(Eq. 4)}$$

x_dev = x – mean(x);

5.5.4. Variance

The variance coefficient is used to quantify the dispersion of a signal, in other words its tendency to remain away from the mean.

$$\mathrm{var}(x) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \qquad \text{(Eq. 5)}$$

x_var = var(x);

5.5.5. Standard Deviation

By definition, the standard deviation is the square root of the variance and is traditionally designated by the Greek letter σ. The standard deviation is also used to quantify the dispersion of a signal but, thanks to the addition of a square root, has the same measurement unit as the signal itself. This is why it is so popular: σ can be directly compared to the signal itself.

$$\sigma(x) = \sqrt{\mathrm{var}(x)} \qquad \text{(Eq. 6)}$$

x_sigma = std(x);

5.5.6. Moment

The m^{th} moment of a signal is defined as the average of the m^{th} power of all signal values. Of more practical usefulness is the central moment, i.e. the moment of a mean removed signal. For a discreet signal, the central moment is defined as:

$$M^m = E((x - \bar{x})^m) \qquad \text{(Eq. 7)}$$

x_m = moment(x,m);

where M^m is the moment of order m. The central moment is extremely useful because it permits to generalise the concept of variance: We can now see that the variance is the second order central moment. The first order central moment is zero.

Moments of higher order than the second are not often used, except for the calculation of skewness (see section 5.5.7) and Kurtosis (see section 5.5.8). It is however worth noting two things: 1) If the PDF is symmetrical in form, the third and higher odd-order central moments must be zero, and therefore such moments give an indication of asymmetry (or "skew") in a distribution. 2) Moments of higher order pay increasing attention to the extreme values of a signal.

5.5.7. Skewness

The skewness is a measure of the asymmetry of the data around the sample mean. If the skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. The skewness of the normal distribution[2] (or any perfectly symmetric distribution) is zero. The skewness of a distribution is defined as:

$$S = \frac{M^3}{\sigma^3} \qquad \text{(Eq. 8)}$$

x_skew = skewness(x);

5.5.8. Kurtosis

The kurtosis is a measure of how outlier-prone a distribution is. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3.

The kurtosis of a distribution is defined as:

---

[2] The normal distribution is also termed Gaussian distribution. It is aa symmetric distribution that is uniquely defined by its mean and standard deviation.

$$K = \frac{M^4}{\sigma^4} \qquad \text{(Eq. 9)}$$

x_kurt = kurtosis(x);

5.5.9. Covariance

The covariance coefficient (or covariance product) is used to question the similarity of two signals. The meaning of the covariance coefficient can consequently be analysed. A zero value means that one signal is pure random noise.

$$C(x,y) = E[(x-\mu_x)(y-\mu_y)] \qquad \text{(Eq. 10)}$$

xy_cov = cov(x,y)

The cross-covariance sequence is a signal constituted of covariance coefficients obtained by time shifting one of the signals. In other words, this represents an extension to the concept of covariance where we look for interdependence of two signals by sliding the signals along one another prior to calculating their covariance. In the equation below, the symbol m denotes the number of samples from which the signal y has been shifted.

$$\Phi(x,y,m) = E[(x_n - \mu_x)(y_{n+m} - \mu_y)] \qquad \text{(Eq. 11)}$$

xy_cov_seq = xcov(x,y,maxlag);
%maxlag is the maximum value of m

The auto-covariance sequence is the same as the cross-covariance coefficient with the exception that both signals are identical. Typically, the auto-covariance sequence might help in telling us about the hidden periodicities of an apparently noisy signal: a periodic auto-covariance sequence indicates the presence of periodic components in the signal whereas a flat auto-covariance sequence means that the signal under investigation is truly non-periodic.

5.5.10. Correlation

The correlation coefficient (or correlation product) is a normalised covariance coefficient whose values are delimited by the range [–1; + 1]. The absolute value of the correlation coefficient indicates the degree or resemblance of the signals x and y whereas its sign indicates the phase of x with respect to y. The correlation coefficient has this advantage over the covariance coefficient in expressing the resemblance of the signals x and y, regardless of their respective amplitude ranges.

$$\rho(x,y) = \frac{C(x,y)}{\sqrt{C(x,x,) * C(y,y)}} \qquad \text{(Eq. 12)}$$

xy_cor = corrcoef(x,y);

Note that √[C(x,x) *C(y,y)] can also be expressed as √[var(x)*var(y)], or simply σ(x)*σ(y).
A signal made of correlation coefficients, called the correlation sequence, might be obtained in the same fashion as the covariance sequence, by sliding the two signals against one another. The cross-correlation sequence is the correlation sequence of two different signals and the auto-correlation sequence is the correlation sequence of the same signal.

$$\rho(x,y,m) = \rho(x_n, y_{n+m}) \qquad \text{(Eq. 13)}$$

xy_cor_seq = xcorr(x,y,maxlag,option);
%maxlag is the maximum value of m *and option is the*
%sequence normalising process.

Because signals x and y contain an identical finite number of samples N, their correlation sequence will become affected by the "end effect". This effect happens when the signals being correlated no longer overlap to form pair products at the beginning and the end of the correlation sequence. As a result, the correlation sequence exhibits a maximum near $m = 0$ and decreases in a linear fashion on each side of the y-axis. This linear decrease can be reduced by making one signal twice as long as the other and limiting the range of the shifting variable m (to maximise the number of pair products in each correlation window) or by employing appropriate normalisation processes. The normalisation processes described below are the ones currently used by the Matlab correlation function *xcorr*. The raw correlation sequence corresponds to σ(x,y,m). The biased correlation sequence is expressed as:

$$\rho(x,y,m)_{biased} = \frac{1}{N}\rho(x,y,m) \qquad \text{(Eq. 14)}$$

xy_cor_seq = xcorr(x,y,maxlag,'biased');

The biased correction does not limit the end effect but provides normalisation to the amplitude range scale of the correlation sequence. The unbiased correlation sequence is expressed as:

$$\rho(x,y,m)_{unbiased} = \frac{1}{N-|m|}\rho(x,y,m) \qquad \text{(Eq. 15)}$$

xy_cor_seq = xcorr(x,y,maxlag,'unbiased');

The normalising factor 1 / (N - |m|) cancels out the linear decrease and normalises the amplitude range of the correlation sequence. This normalisation process is elegant because it hides the loss common terms in the correlation. However, we must remember that the resulting correlation sequence still has its maximal accuracy near m = 0.

5.5.11. Convolution

Like correlation and covariance, convolution of digital signals is a sum of product type of operation. However, it differs from the two previous by requiring that one of the variables must be time inverted (or one could say horizontally flipped). Although convolution is of no interest in deriving statistical properties from a signal, it is included here because of its similarity of process with the correlation product.

Convolution, however, is the mathematical operation behind linear filtering. A signal x is said to be filtered when it has been convoluted with the impulse response h of a filter. The impulse response is a sequence of values that carries the filter characteristics and it corresponds to the inverse Fourier transform of the spectral characteristics of the filter.

$$x_n \otimes h = \sum_{i=0}^{M-1} x_{n-i} \cdot h_i \qquad \text{(Eq. 16)}$$

```
y(n) = sum(x(n-M+1:n).*h(1:M)); % for a single value n
y = conv(h,x); %for an entire signal
y = filter(h,1,x); %for an entire signal
```

The convolution product has its specific notation ⊗ to distinguish it from the conventional product. The equation above shows the convolution product of the impulse response h with the input signal x at moment n. The length of the sum of products is M values and this corresponds to the length of the impulse response. As M is a finite number, the filter represented by the equation above is termed a Finite Impulse Response (FIR) filter. FIR filters have the characteristics of being non-recursive, i.e. their output value at moment n depends only on the past and present input values, and not on the past output values. To obtain an intuitive understanding of the convolution product, we must simply regard it as weighted average of the last M values of a signal, starting at moment n. By computing this weighted sum every incremental step of n, we produce a set of values that constitute the filtered signal.

5.6. Frequency domain transform

5.6.1. Fourier Transform

We primarily experience a signal in the time domain. However, the knowledge of its spectral signature is very important as it tells us about the harmonic components (or sub signals) that populate this signal. The Fourier Transform (FT) is employed to cross the barrier between time and frequency representations of that signal. The two domains provide complementary information about the same data. Below is the Fourier integral that produces the FT of a time signal of infinite duration.

The FT is a particular case of the Fourier series when the signal to be analysed is not necessarily periodic.

$$X(f) = F[x(t)] = \int_{-\infty}^{\infty} x(t) \cdot e^{-j2\pi ft} dt \qquad \text{(Eq. 17)}$$

The term j represents the complex variable of value $\sqrt{(-1)}$.

The squared modulus of X(f) is known as the periodogram, that is to say an estimation of the Power Spectral Density (PSD) of x(t). Assuming x(t) has unit V, the periodogram has unit V^2/Hz .

In comparison with the periodogram, the true PSD is defined as the DFT of the autocorrelation function of x(t). The periodogram yields almost similar results, but without the need for calculating the autocorrelation sequence, therefore speeding up the process.

The periodogram method, as well as other spectral analysis methods, are more specifically discussed in section 5.11.

5.6.2. Inverse Fourier Transform

The FT has numerous mathematical properties, including linearity and invariance through time translation. Also, the Fourier transform is reversible: a time domain signal can be obtained from a frequency signal by applying the Inverse Fourier transform (IFT).

$$x(t) = F^{-1}[X(f)] = \int_{-\infty}^{\infty} X(f) \cdot e^{+j2\pi ft} df \qquad \text{(Eq. 18)}$$

5.6.3. Discrete Fourier Transform

The FT was initially developed for continuous infinite length real or complex valued signals. If it is to be calculated by a digital computer, it must be adapted for discrete time finite length signals. This adaptation is called the Discrete Fourier Transform (DFT) and is given below.

$$X(k) = \sum_{0}^{N-1} x(n) \cdot e^{-j2\pi k\Omega nT} \qquad \text{(Eq. 19)}$$

No Matlab equivalent syntax since the DFT is always calculated by FFT. See section 5.7.

The passage from continuous to discrete time produces a discretisation of the number of frequency values that can be resolved with the DFT. In other words, the frequency variable f is no longer continuous but discrete, and this implies there is a finite frequency interval between two consecutive values of the discrete f. This interval is inversely proportional to the duration of the sampling sequence.

In the same manner that time sampling replaced t by nT, we now replace f by kΩ, k being the integer frequency index and Ω, the smallest frequency interval. If a harmonic has a frequency kΩ, (whatever the value of k) then the harmonic is perfectly resolved. However, for all cases where the harmonic has a frequency represented by (k+0.5)Ω, then the energy of this harmonic is shared in the spectrum between the two upper and lower neighbour frequencies kΩ and (k+1)Ω. In other words, the energy associated with this harmonic is not sharply represented but spread out onto its neighbours. This effect is called leakage. To minimise leakage, one must make sure that the sampling frequency is high enough to permit a correct resolution of the signal's harmonics and also use data set of the longest possible duration to decrease the frequency step Ω.

The DFT assumes that x(nT) is periodic with a period NT. This is clearly not the case at all times and special consideration must be given to x so that x(0) is not too different from x((N-1)T). Failure to do so results in the distortion of the spectrum because of the periodic discontinuity produced by the junction of x(0) to x((N-1)T). To ensure that these two quantities do not produce discontinuities, they can be made equal to zero by windowing the signal, as we will see in section 5.11 on frequency domain signal analysis.

If one is familiar with the concept of correlation, the mechanism of the DFT can be intuitively explained. First, one must recognise that the function $e^{-j2\pi k\Omega nT}$ is a complex harmonic signal generator. By progressively incrementing the frequency f and computing the FT, we obtain a signal X(f) that contain information about the amount of the harmonic in x(t) at frequency f. In other words, the signal x(t) is being "scanned" for the presence of harmonics, f being the scanning variable. The detection of a particular harmonic in x(t) is simply a correlation process where x(t) is correlated with the harmonic signal generator $e^{-j2\pi k\Omega nT}$.

Like the FT, The DFT is reversible: a time domain digital signal can be obtained from a frequency domain digital signal by applying the Inverse Discrete Fourier transform (IDFT).

$$x(n) = \sum_{0}^{N-1} X(k\Omega) \cdot e^{+jk\Omega nT} \qquad \text{(Eq. 20)}$$

5.7. Fast Fourier Transform

A large number of multiplication's and additions are required for the computation of the DFT and its direct computation is an intensive process that penalises its usefulness. Part of the computational load comes from the redundancies produced by the n*k product in the harmonic signal generator [Ifeachor and Jervis (1993)]. The Fast Fourier Transform (FFT) algorithm uses computational recipes to eliminate the computational complexity of the DFT by eliminating these redundancies. Since 1965 when it was introduced by Cooley and Tukey, the FFT has gained an enormous importance from providing acceleration to the DFT computation. [Bellanger (1994)] In Matlab notation, the FFT of a signal x and IFFT are written:

```
X = fft(x);
x = ifft(X);
```
(Eq. 21)

5.8. Fast convolution with the FFT

The convolution theorem states the equivalence between the Fourier transform of a convolution product of two signals and the simple product of the Fourier transform of each signal. This reads:

$$F(x(t) \otimes y(t)) = X(f) \cdot Y(f) \qquad \text{(Eq. 22)}$$

The convolution product is found by applying an inverse Fourier transform to both sides of the previous equation.

$$x(t) \otimes y(t) = F^{-1}(X(f) \cdot Y(f)) \qquad \text{(Eq. 23)}$$

In principle, a convolution can be realised with identical results either in the time or frequency domains, thanks to the application of the convolution theorem. This is also valid for discrete signals and therefore the Fourier transform can be replaced by the FFT. When the signals contain a large number of points (>1000), it is the case that the convolution product becomes more quickly calculated in the frequency domain than in the time domain, thanks to the computational rapidity of the FFT.

Paradoxically, the convolution is often calculated in the time domain. This is for many reasons: Firstly, convolution is the operation behind linear filtering (see section 5.14) and, when filtering low frequency data, time domain calculation yields more accurate results than FFT-based calculation. This is because the frequency resolution of the FFT is limited by the duration of the data set. Secondly, real-time filtering is not possible with FFT-based convolution. Thirdly, the argument that FFT-based filtering is faster than time time-domain filtering is progressively loosing importance given the rapidity of modern computers.

The Matlab method for computing the convolution are described in section 5.5.11.

5.9. Fast correlation with the FFT

Similarly to convolution, we can perform FFT based cross and autocorrelation (with the same advantages in computational rapidity) thanks to the correlation theorem. The cross-correlation equation is given below:

$$\rho(x, y, m) = F^{-1}[F[x(n)] \cdot F[y(n+m)]^*] \qquad \text{(Eq. 24)}$$

xcorr is an FFT based correlation process, see equation 13 in section 5.5.10.

The * symbol denotes the conjugate of a complex number. Taking the complex conjugate of the Fourier transform of a signal is equivalent to time inverting it (or one might say flipping it horizontally). To perform an auto-correlation, it is only necessary to replace y by x. As in the case of the convolution, this operation is valid for discrete signals and therefore FT and IFT can be replaced by FFT and IFFT.

5.10. The Parseval relation

The Parseval relation states that the power of a time signal is equal to the sum of the power of its harmonics. In other words, there is a time-frequency equivalence in the total energy carried by a signal. This is represented below mathematically as:

$$\frac{1}{N}\sum_{n=0}^{N-1}|x(n)|^2 = \sum_{k=0}^{N-1}|X(k)|^2 \qquad \text{(Eq. 25)}$$

5.11. Classical spectral analysis with the FFT

Spectral analysis is a large and much research subject, but it is most often defined as an estimation of the PSD. This estimation is commonly calculated with the periodogram method, which is the workhorse of spectral analysis, and the most ancient method since the introduction of the FFT. Section 5.11 will present the periodogram (already introduced in section 5.6.1) and its variants, as well as the Blackman-Tukey method. The latest is also a method for estimating the PSD, but not a periodogram.

All these methods fall into the "classical" category because they are FFT-based. This is not necessarily the case in modern spectral analysis methods. Modern spectral analysis is outside the scope of this thesis because only classical methods were used on the FORP signals.

The periodograms of three different TCG signals are represented in section 12.4.1.3, figures 6 to 8.

The Power Spectral Density (PSD) is defined as the FT of the autocorrelation function of a signal [Lynn (1982)].In discreet form, this is expressed as:

$$PSD(f) = \sum_{m=-\infty}^{\infty} c_{xx}(m) \cdot e^{-j2\pi fm} \qquad \text{(Eq. 26)}$$

where f is the frequency in Hz, and c_{xx} is the autocorrelation function. Thanks to its averaging effect, the autocorrelation function only contains very little random noise. Also, the autocorrelation effect suddenly makes the periodicities in the signal very apparent. This is why taking the FT of the autocorrelation yields a good spectral representation of the signal.

However, the determination of the PSD using the equation above is problematic because it requires an infinite set of autocorrelation coefficients.

It is important to know that the PSD was introduced before the FFT algorithm. We can now understand why the direct calculation of the PSD was traditionally avoided: The burden of computing the DFT on top of the autocorrelation function must have been discouraging.

5.11.1. The periodogram

The periodogram method for estimating the PSD, which was also introduced prior to the FFT, came as an answer to the complexity of the direct computation of the PSD. Suddenly, the PSD could be estimated without the need for autocorrelating the signal: The periodogram is the squared modulus of the DFT of a finite signal. This is expressed as:

$$PSD_{est}(f) = \left| \sum_{n=0}^{N-1} x(n) \cdot e^{-j2\pi fnT} \right|^2 \qquad \text{(Eq. 27)}$$

The periodogram was the simplest spectral analysis method, and is still largely used. Naturally, its popularity even further increased after the introduction of the FFT in 1965.

This method has the double advantage of its simplicity and computational rapidity. However, the periodogram is a biased estimator of the true PSD. The estimation is inconsistent because its variance (from successive repetitions of the same experience) is very large [Ifeachor and Jervis (1993)].

The periodogram is often called power spectrum. The square root of the periodogram is also called amplitude spectral distribution, or simply spectrum.

5.11.2. The modified periodogram

We have seen in section 5.6.2 that the DFT assumes that the discrete signal which is transformed to the frequency domain is periodic of period equal to the length of this signal. To limit the effect of this artefact periodicity, one can "window" the time signal prior to taking its DFT. The multiplication of the window to the time signal is intended to create an artificial continuity between the beginning and the end of the windowed time signal by forcing both ends to zero values. A typical tapering window has the same duration as the signal to be analysed and possess a maximum in the middle. The window is symmetrically shaped and the specificity of this shape defines its frequency characteristics. Because the window is multiplied by the time signal, we can also say (thanks to the convolution theorem) that the signal's spectrum is convolved with the window's spectrum.

In the periodogram method, the direct application of the FFT on a discrete time signal with a finite number of samples can be regarded as a windowing operation where an infinite duration signal is multiplied by a rectangular window (also called a gate or boxcar window). In other terms, rectangular windowing (or gating) is inherent to the application of the FFT on an infinite duration time signal.

The FT of a time gate is called a Dirichlet kernel, or cardinal sine function (or sinc function), and is expressed as $(\sin x)/x$, where x is the frequency variable. The effect of time gating on a spectrum obtained from a discrete signal of infinite duration is to convolve the signal's FT with a Dirichlet kernel. This leads to an estimated spectrum being a low pass version of the true spectrum because the cardinal sine function can be regarded as the impulse response of a low pass filter. The cardinal sine function has one large central lobe and multiple side lobes. The central lobe provides the low

pass effect, also termed smearing. The side lobes produce a phenomenon called spectral leakage[3]: The side lobes are responsible for spreading the energy density of an harmonic all around its central frequency, thus producing a "leak". The leakage can be reduced by carefully choosing a window with little as side lobes as possible. There are numerous window types, each with a specific effect on the spectrum estimation but they all have in common the smoothing of the spectrum in order to remove randomness [Bendat and Piersol (1986)].

The modified periodogram is a windowed periodogram. This is expressed as:

$$PSD_{est}(f) = \left| \sum_{n=0}^{N-1} w(n) \cdot x(n) \cdot e^{-j2\pi fnT} \right|^2 \qquad \text{(Eq. 28)}$$

where w(n) is the time window.

5.11.3. The Bartlett modified Periodogram

In the Bartlett modified periodogram, to yield more accurate and consistent results and to reduce the variance, the signal is divided into K non-overlapping sections of length M (with M<N) and the periodogram of each segment is taken. The average of all K periodograms is taken to be a better estimation of the PSD.

Sectioning the data results in fewer samples per FFT and consequently a reduction in frequency resolution, but this may be overcome by the addition of augmenting zeros at the end of each subsection. However, averaging periodograms coming from the same signal reduces the variance.

It is necessary to bear in mind the opposing claims of the accuracy of the estimate and the required frequency resolution, and to look for the best compromise.

The Bartlett modified periodogram is expressed as:

$$PSD_{est}(f) = \left| \frac{1}{K} \sum_{k=0}^{K-1} \sum_{n=kN/K}^{(k+1)(N/K)-1} w(n) \cdot x(n) \cdot e^{-j2\pi fnT} \right|^2 \qquad \text{(Eq. 29)}$$

---

3 We have previously explained leakage in section 5.6.2 as the impossibility of perfectly resolving harmonics whose frequency is not an integer multiple of the frequency resolution, resulting in the harmonics power being shared with the surrounding frequency bands.

5.11.4. The Welch Modified Periodogram

Welch proposed an alternative to the modified Bartlett periodogram by allowing the data segment to overlap. Consequently the size of each signal section can be increased without "zero padding" the sections and it can be shown that the final spectrum variance is further decreased compared to the Bartlett periodogram [Welch (1967)]. This method is designated as the Welch modified periodogram, and corresponds to a generalisation of the Bartlett modified periodogram.

5.11.5. The Blackman-Tukey method

The methods of spectral estimation described previously (periodogram, Bartlett and Welch modified periodograms) have fundamental limitations: namely, tapering of windows which is equivalent to low pass filtering the spectral waveform, hence a lost of frequency resolution. This tapering while on the one hand giving benefits in terms of reduction of spectral leakage and variance, can, on the other hand, lead to suppression of artefacts and features in spectrum that may be of considerable interest.

An alternative method proposed by Blackman and Tukey is to generate an estimation of the PSD by windowing the autocorrelation function and taking the absolute value of the FFT of the whole. This method was introduced in 1958, seven years before the FFT. In its form, it is obviously a much closer estimation of the PSD than the periodogram, but it only became popular when the FFT permitted fast calculation of the autocorrelation function (see section 5.9).

The PSD estimation by the Blackman-Tukey method is:

$$PDF_{est}(f) = \sum_{m=-M+1}^{M-1} r_{xx}(m) \cdot w(m) \cdot e^{-j2\pi fm} \qquad \text{(Eq. 30)}$$

By comparison with the Welch and Bartlett periodogram methods, we see that the smoothing is achieved by the averaging effect of the autocorrelation process rather than by the averaging of several periodograms.

The autocorrelation is windowed to taper it towards its extremes because at larger lags fewer data points enter the computation so these estimates are less accurate. Tapering has the effect of attaching less weight to these estimates.

5.12. Laplace Transform

The Laplace transform (LT) is the equivalent of the FT for continuous causal signals, (i.e. analogue signals for which values are only known within the time boundaries of the acquisition experiment). The Laplace integral can be regarded as a special case of the Fourier integral where negative time is not allowed because the experiment consisting of acquiring the signal s(t) starts at time t=0. Also, the Laplace transform has a specific variable usually called p that is defined as $p = \sigma + j\omega$, where $j\omega$ is the Fourier frequency variable.

$$L[x(t)] = \int_0^\infty x(t) \cdot e^{-pt} dt = X(p) \qquad \text{(Eq. 31)}$$

5.13. Z Transform

The Z transform (ZT) was invented to permit a simple description of discrete causal signals and systems (such as digital filters). Like the FT, the ZT can reduce convolution products into simple products by use of the convolution theorem. Unlike the FT, the ZT cannot be used for spectral analysis directly. The ZT is defined below.

$$Z[x(n)] = \sum_{n=0}^{\infty} x(n) \cdot z^{-n} = X(z) \qquad \text{(Eq. 32)}$$

This actually is a power series and therefore will not always converge. Also, we can see immediately that, in the case where $z = e^{jk\Omega T}$, the ZT is equivalent to a DFT. Like the FT, the ZT is a linear transformation (the ZT of a weighted sum is the sum of weighted ZTs). Also, the ZT of a convolution product is the product of the ZTs.

The ZT of a digital system is usually expressed as a ratio of two polynomials. Let us consider the example of a mobile average filter. Its Finite Impulse Response (FIR) vector is a time gate so the impulse response has value 1 in the range [0:M-1] and 0 anywhere else. Consequently, the ZT of the impulse response is given by:

$$Z(h(n)) = \sum_{n=0}^{M-1} h(n) \cdot z^{-n} = 1 + z^{-1} + z^{-2} + z^{-3} + \ldots + z^{-M+1} = H(z) \qquad \text{(Eq. 33)}$$

In digital filtering, almost any process requires the knowledge of the history of a signal and therefore time shifting is an elementary operation. By looking carefully at the equation above, we understand that the z variable let us express the time shift of a discrete signal easily: In the z domain, to delay a signal by k samples it is only sufficient to multiply it by z^{-k}. We must remember that expressing the same delay analytically in the Fourier domain requires multiplying the signal's

FT by a complex exponential term. This is obviously a longer and more intensive operation. We now clearly see the enormous advantage of the ZT over the FT in DSP design, as it allows us to symbolically manipulate time delays easily.

The power series described in the equation above, which is called the transmittance in Z of the filter, is generalised as:

$$H(z) = \frac{\sum_{i=1}^{B} b_i \cdot z^{-i-1}}{1 - \sum_{i=1}^{A} a_i \cdot z^{-i}} = \frac{N(z)}{D(z)} \qquad \text{(Eq. 34)}$$

where a is the coefficients vector of the denominator D(z) and b is the coefficients vector of the numerator N(z). In our example of mobile average, the impulse response h is equal to the vector b and the vector a is zero[4] so we could write H(z) = N(z). The roots of polynomial N(z) are called the zeros and the roots of polynomial D(z) are called the poles. Zeros cancel out the transmitance at particular values of z (hence for particular frequencies) whereas poles cancel out D(z) and make the transmittance diverge. Poles and zeros, like the variable z, are complex numbers represented on the z- plane. To find the poles and zeros of a digital system, it necessary to factorise N(z) and D(z). This is easy for polynomials of order ≤ 2 but order 3 and above are generally solved with the help of a numerical method.

5.13.1. Stability of a digital filter

The ZT allows us to easily address an important question in filter design i.e. the stability of the filter. A stable filter will only react in the presence of an input and the impulse response will always tend to zero after a certain time. If the filter is unstable, it might produce an output without input, in the manner of a random noise generator or an oscillator. However we certainly do not want our filters to behave as oscillators and for this matter we must examine the conditions that make the filter transmittance stable. Finite Impulse Response filters are naturally stable because their D(z) is always equal to 1 (see previous equation) and consequently H(z) cannot diverge. The position of zeros in the z plane does not affect the stability of the transmittance. Infinite Impulse Response filters are stable only if their poles are inside the unit circle in the z plane. This can be proven by showing that poles with absolute values greater than 1 produce diverging impulse responses, hence resulting in unstable systems. Representation of poles and zeros in the Z plane allows us to quickly and efficiently assess the stability of a digital filter.

---

4 The vector a contains the recursive coefficients of an Infinite Impulse Response (IIR) filter. See section 5.14.2.

5.13.2. Frequency response analysis

The positions of poles and zeros in the Z plane also provides us with an intuitive idea of the frequency response of the system. It is important to note that a zero in the z domain only means an attenuation of the frequency response, except if the zero lies on the unity circle in which case the frequency response is null. By changing z into the complex exponential term $e^{jk\Omega T}$, the transmittance H(z) becomes the frequency response H(j kΩ), i.e. the exact frequency response that would have been obtained by DFT. The Matlab function "freqz" permits to realise this transformation.

5.13.3. Input output relation

In ZT terms, the filter's input/output relation is given by:

$$Y(z) = H(z) \cdot X(z) \qquad \text{(Eq. 35)}$$

where X(z) and Y(z) represent, respectively, the Z transform of x(n) and y(n). Knowing the coefficients of a digital filter lead immediately to the knowledge of H(z). Refer to equation 34.

5.13.4. Inverse Z transform

Like the DFT, the ZT is a reversible transform and the Inverse ZT (IZT) is denoted symbolically as:

$$h(n) = Z^{-1}[H(z)] \qquad \text{(Eq. 36)}$$

The actual process of inverse transforming is not reversible and must be obtained by one of the three following methods: Power series expansion, partial fraction expansion, or a residue method [Bellanger (1994)]. However, most common elementary time sequences (Dirac function, step function, sinc function, linear slope function) have a tabulated ZT and this helps in rapidly deriving the IZT. [Reynaud (1995)]

5.14. Linear filtering

This section has been left to last as it employs all the materials previously covered in this chapter. A filter is initially a physical system that modifies the spectral content of a signal. A linear filter, or simply filter in the following, is a process operating a convolution product i.e. a sum of linear products between a section of the input signal and the impulse response of the filter.

The filter has its own spectral signature (or transmittance) and temporal signature (or impulse response). The impulse response is the filter's response to a Dirac input. The transmittance is the Fourier transform of the impulse response. To apply a filter to a signal, the impulse response is

simply convolved with this signal. Thanks to the convolution theorem, this operation can also be done in the frequency domain to make it quicker by taking advantage of the computational efficiency of the FFT algorithm.

Filters are usually derived from physical systems, such as analogue RC filters for example. The use of a digital computer allows us to simulate any analogue filters. The power of DSP compared to analogue processing is that there is no need to be realistic in the simulation of a real system. This permits us to derive filters that could not be implemented physically with analogue electronics. For example, a moving average is a non-recursive filter that has no analogue equivalent because it was actually derived from the statistical concept of mean value. We can distinguish two types of linear filters: Finite Impulse Response (FIR) filters and Infinite Impulse Response filters. Their specific properties are reviewed below.

5.14.1. Finite Impulse Response

The FIR filters have a finite duration impulse response. This is due to the absence of recursivity in their algorithm. All physical systems have recursivity and that makes FIR typically specific to the digital domain. The input/output relation of time domain signals of a causal FIR filter is expressed below:

$$y(n) = \sum_{i=0}^{M-1} h(i) \cdot x(n-i) \qquad \text{(Eq. 37)}$$

y = filter(h,1,x);

with h being the FIR vector of the filter. The FIR vector is a sequence of coefficients. It can be regarded as a specific wave shape to be identified within the input signal. In a process that resembles a correlation product, the magnitude of the convolved signal varies with the degree of resemblance between the signal section currently analysed and the wave shape of the FIR vector h[5].

FIR filters have the very interesting property that, for periodic input signals, the input/output delay (or phase delay) is constant and predictable when the impulse response is symmetrical. This is not demonstrated here but at the end of appendix 2 on phase cancellation filter. This is an immense advantage when designing filters for real-time operation. In the case of the filter described above whose FIR vector contain an odd number of values M, the phase delay would be (M-1)/(2Fs), with Fs being the system sampling frequency. This constant phase delay provides a linear phase variation against frequency between input and output signals. This is unlike IIR filters whose phase variation is non-linear. Linear phase FIR filters respect the phase delay structure between all the

5 This intuitive approach to understanding FIR filtering is the basis of "matched filtering".

harmonics of the input signal. One can therefore say that they introduce no time distortion. [Lynn 1982]

The phase delay of a linear phase FIR filter can even be reduced to precisely zero if one is running an offline processing. This is demonstrated at the beginning of appendix 2. The equation given below shows a non-causal convolution since the convolution product boundaries extend into the future [Press et al. (1992)]. Of course, a zero delay means also a zero phase.

$$y(n) = \sum_{i=-M/2}^{M/2-1} h(i + M/2) \cdot x(n - i) \qquad \text{(Eq. 38)}$$

```
y(n) = sum(h(1:M).*x(n-M/2:n+M/2) );
```

The zero phase property of the symmetrical convolution product is simply explained: by shifting the impulse response so that it covers the past and the future in equal proportions, the calculation of the convolution product output value (situated in the present) sees the contribution of exactly the same number of input values of the past and the future, weighted in the same fasion. The symmetry of the convolution product only affects the phase of the output signal by cancelling it but does not modify its wave shape, compared to the causal convolution product. Symmetrical convolution with a linear phase FIR filter is a great tool for offline signal filtering and analysis because it leaves the phase structure of the input signal absolutely intact.

The filter's behaviour is completely described by the coefficients contained in the FIR vector h. The purpose of filter design is to generate a set of coefficients that will adequately attenuate or amplify particular regions of the input signal's spectrum.

Methods for deriving the coefficients are numerous. The three most common methods for calculating the filter coefficients are (1) the window method, (2) the frequency sampling method, (3) the optimal method.

5.14.1.1. The window method

The window method is essentially an analytical method that produces a FIR vector by calculating the inverse Fourier transform of an ideal frequency response. (Usually, such transform is not calculated but obtained from a table.) For example, a perfect low pass filter would consist of a gate in the frequency domain. The gate's IFT is a Dirichlet kernel, also commonly called cardinal sine function, or "sinc". After sampling the Dirichlet kernel to discretise it and truncate it to a finite duration, the resulting sequence can be used as a FIR vector directly. However, it is recommended that it is multiplied by an appropriate tapering window in order to reduce the discontinuities arising from truncation. See figure 2 on the next page.

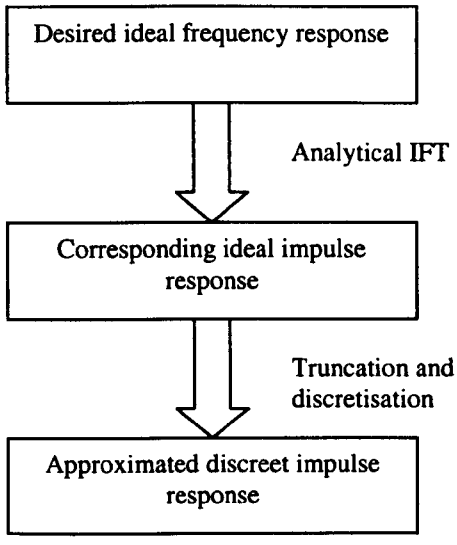

Figure 2: The Window method for FIR filter design

This method is the simplest of all and probably the most intuitive. However, if the sampling parameters (duration, sampling frequency, and number of samples) are too low, the approximated FIR vector might not coincide very well with the desired frequency response. In particular, the frequency response corresponding to the discretised impulse response will contain unwanted ripples of variable amplitude. This method is not recommended for real time operation but can give superb results when efficiency of calculation time (due to carrying out the convolution process for a long FIR vector) is not critical.

5.14.1.2. The frequency sampling method

The frequency sampling method uses a similar approach to the window method with the difference that the desired spectral response is sampled and truncated directly in the frequency domain and its IDFT is taken to obtain a FIR vector. See figure 3. This numerical method allows us to produce FIR vectors from any frequency responses. By comparison, the window method can only be applied to frequency responses that can be inverse Fourier transformed analytically.

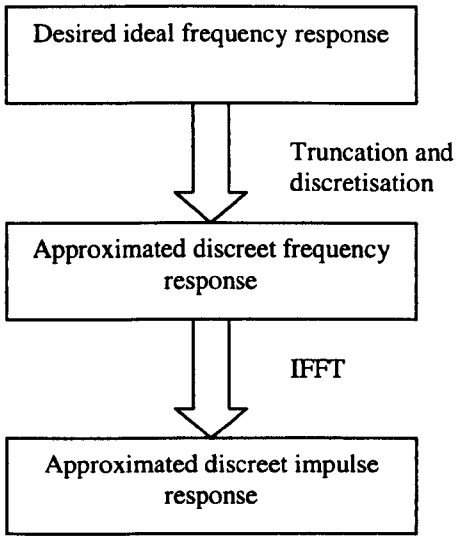

Figure 3: The Frequency sampling method for FIR filter design

5.14.1.3. The optimal method

The optimal method for FIR vector coefficient calculation is based on the minimisation of the maximum magnitude difference between the desired frequency response (i.e. a frequency response based on frequency domain gating) and the real frequency response. This is mathematically denoted below:

$$Min[Max[E(jw)]] \qquad \text{(Eq. 39)}$$

where E(jω) represents the difference between the real and the desired frequency response. It has been proved [Rabiner and Gold 1975] that when max[|E(jω)|] is minimised, the filter frequency response will show ripples of equal amplitude (or equi-ripple) regardless of the concerned frequency band.

Additionally, it must noted that SPTOOL, the graphic user interface for signal processing in Matlab provides ready to use filter design algorithms and time-frequency analysis tools.

5.14.2. Infinite Impulse Response filter

From an algorithmical point of view, IIR filtering is an extension of FIR filtering by introduction of a recursive term in the filter equation. Historically however, IIR filtering has been the only type of filter available until the invention of signal processing with digital computers. The input/output relation of time domain signals of a causal IIR filter is expressed below:

$$y(n) = \sum_{i=0}^{B-1} b(i) \cdot x(n-i) + \sum_{i=1}^{A} a(i) \cdot y(n-i) \qquad \text{(Eq. 40)}$$

```
y = filter(b,a,x);
```

IIR filters are based on two independent, finite duration convolution products. The first convolution term represent the non-recursive part of the filter and the second term is the recursive part. The vector b contains the non-recursive coefficients and the vector a holds the recursive coefficients. In comparison to FIR filters, IIR filters seem more computationally intensive because of the two independent convolution products. However, one of the properties of recursivity is that it helps to reduce the number of coefficients required (and hence the number of mathematical operations) to generate a filtering effect that would be equivalent to a FIR filtering effect.

IIR filters have a non-linear phase response due to the varying phase delay at each frequency. For that reason, IIR filters are not used in applications where no phase distortion is required.

By inspection of the equation above, the term Infinite Impulse Response does not seem to be justified since both convolution products involve finite duration coefficient sequences. However, if we attempt to re-write the preceding equation without the recursive term, one rapidly concludes that the positive boundary of the non-recursive convolution term is infinite, therefore justifying the term IIR.

By taking the ZT of the IIR filter equation above, we obtain:

$$Y(z) = \left[X(z) \cdot \sum_{i=0}^{B-1} b(i) \cdot z^{-i} \right] + \left[Y(z) \cdot \sum_{i=1}^{A} a(i) \cdot z^{-i} \right] \qquad \text{(Eq. 41)}$$

where a and b are, respectively, the recursive filter coefficients vector and the non-recursive coefficients vector. We define the ratio of Y(z) / X(z) as the transmittance in Z of the IIR filter. H(z) is given by:

$$H(z) = \frac{\sum_{i=0}^{B-1} b(i) \cdot z^{-i}}{1 - \sum_{i=1}^{A} a(i) \cdot z^{-i}} \qquad \text{(Eq. 42)}$$

They are numerous ways to derive the coefficients of an IIR filter. The simplest is the zero-pole placement method, where an IIR filter is produced intuitively by "dropping" appropriate zeros and pole on the z plane. However the leading concept in IIR filter design is to simulate real analogue filters, like Butterworth or Chebyshev filters for example. The main advantage of this approach is

that the resulting digital filter is modelled on a filter whose characteristics are exactly known in advance. The second advantage is that it only requires analytical work to derive the filter's coefficients. Methods based on pre-existing analogue filters are the impulse invariant method, the filter equation method, and the bilinear ZT method. All of these methods are reviewed below.

5.14.2.1. Zero-pole placement method

This method is the inverse approach of the common z plane based frequency response analysis method.

When analysing a digital filter, whose transmittance in Z is known, we can derive its frequency response by studying the placement of its poles and zeros. Inversely, in the zero-pole placement method, we establish the frequency response of a system by positioning manually its poles and zeros.

This is a "quick and dirty" method and does not yields precise control of the frequency response.

When a zero is placed at a given point on the z-plane, the frequency response will be attenuated at the corresponding frequency given by the phase of the zero. A pole on the other hand produces a peak. Poles that are close to the unit circle give rise to large peaks, whereas zeros close to or on the circle produce a complete attenuation. Thus, by strategically placing poles and zeros on the z-plane, we can obtain simple low pass, high pass or band pass filters. An important point to bear in mind is that for the coefficients of the filter to be real, the poles and zeros must either be real (that is lie on the positive or negative real axes) or occur in complex conjugate pairs. Finally, the coefficient vectors a and b are obtained by expanding the polynomials N(z) and D(z).

If one is familiar with the significance of pole and zero location in the z plane, this method is by far the simplest and fastest of all. However, it is somewhat restricted to the design of simple filters. If one desires a very specific and sharp frequency response then many poles and zeros will be required and the intuitive understanding of their interactions in the z plane may become difficult.

It must be added that, in theory, this technique can also be used for the design of FIR filters (only zeros, one pole at $z = 0$) but the large numbers of zeros required in a typical FIR filter is a constraint to simple design.

5.14.2.2. Impulse invariant method

This method assumes that we are in possession of a template frequency response equation of the desired filter type, expressed in the Laplace or the Fourier domains. As an example, the template of a second order low pass filter transmittance is expressed below in the Fourier domain:

$$H_{LP}(j\omega) = \frac{H_0}{1 + 2m \cdot \left(\frac{j\omega}{\omega_0}\right) + \left(\frac{j\omega}{\omega_0}\right)^2} \qquad \text{(Eq. 43)}$$

where H_0 is the static gain[6], m is the resonance factor, ω is the angular frequency[7] and ω_0 is the cut off angular frequency. If we can assume that the filter's input signals are periodic then the same transmittance can be expressed in the Laplace domain by replacing the variable $j\omega/\omega_0$ with the variable s. Of course this has the effect of normalising the frequency. Consequently, to use such filter template, the frequency variable s must be denormalised by multiplying it by the desired cut-off frequency ω_0.

The impulse response h(t) of the system is obtained analytically either by ILT or IFT of the frequency response and the transmittance in z (i.e. the digital filter coefficient) is obtained by ZT of the impulse response. Finally, expanding the polynomials N(z) and D(z) of the transmittance in z gives us the filter coefficients.

Most analogue transfer functions, like the one in our example, can be broken down into a sum of first order single pole filters (or simply single pole element). With experience, each single pole element in the Laplace or Fourier transmittance can be expressed as a single pole element in the z domain, thereby bypassing the sequential use of ILT/IFT and ZT. This shortcut greatly simplifies our task.

5.14.2.3. Filter equation method

Like the impulse invariant method, this method assumes a priori knowledge of a suitable analogue filter transmittance $H(j\omega)$. For simplicity, we will assume that this filter corresponds to the second order low pass filter described in the equation above. Hence, $H(j\omega) = H_{LP}(j\omega)$. The input/output relation of such an analogue filter is described as:

$$y(t) = H_{LP}(j\omega) \cdot x(t) \qquad \text{(Eq. 44)}$$

where x(t) is an input continuous real valued signal and y(t) is an output continuous real valued signal. By assuming that the filter's input signal is periodic, we imply that x(t) and y(t) can be completely described as sums of harmonics. We must remember that differentiating harmonics is equivalent to multiplying them by the term $j\omega$. Hence, we can write:

---

6 The static gain is the filter gain for a zero frequency input signal.

7 The angular frequency is defined as $\omega = 2\pi f$ where f is the frequency. The angular frequency is often loosely called the frequency, for simplicity.

$$j\omega \cdot y(t) = \frac{d}{dt} y(t) \qquad \text{(Eq. 45)}$$

which is valid for any periodic signals. By re-arranging the filter input/output relation, we produce a differential equation that is characteristic of the filter's behaviour in the time domain.

$$y(t) + \frac{2m}{\omega_0} \cdot \frac{d}{dt} y(t) + \frac{1}{\omega_0^2} \frac{d^2}{dt^2} y(t) = x(t) \qquad \text{(Eq. 46)}$$

Once the continuous signals x and y and their respective derivatives have been replaced by their discrete equivalent, we obtain an algorithm that can be rearranged to replicate the layout of the IIR filter algorithm. This algorithm is the filter algorithm and is expressed as:

$$a_1 \cdot y_n = (b_1 \cdot x_n) + (a_2 \cdot y_{n-1}) + (a_3 \cdot y_{n-2}) \qquad \text{(Eq. 47)}$$

where n designates the time index and a and b are the coefficients vectors of the digital filter. a and b are expressed below:

$$a_1 = \frac{(\omega_0 T)^2 + 2m\omega_0 T + 1}{H_0} \qquad \text{(Eq. 48)}$$

$$a_2 = 2m\omega_0 T + 2 \qquad \text{(Eq. 49)}$$

$$a_3 = -1 \qquad \text{(Eq. 50)}$$

$$b_1 = (\omega_0 T)^2 \qquad \text{(Eq. 51)}$$

where T is the inverse of the system sampling frequency Fs. Values of vectors a and b can be directly calculated according to their analytical expressions. The expressions for a and b will only change with the filter type (low pass, high pass, order) but not with the filter parameters (cut off frequency, resonance factor, static gain). It follows that a filter of a particular type can be rapidly designed simply by remembering the corresponding expressions for vectors a and b, and providing the adequate filter parameters.

5.14.2.4. Bilinear ZT method

The Bilinear ZT (BZT) method has the same purpose as the impulse invariant method: we aim at deriving the digital transfer function H(z) of a filter by transformation of its analogue transfer function H(s). However both methods differ in the way this is accomplished. The BZT method does rely on single pole decomposition but on a change of variable. To carry out the Laplace to Z domain transformation, we use the following expression:

$$s = \frac{z-1}{z+1} \qquad \text{(Eq. 52)}$$

where s is the Laplace variable and z is the Z variable. This change of variable maps the entire s plane into the z plane. The obvious advantage of this method of deriving H(z) is its simplicity compared to the simple element decomposition required in the impulse invariant method. However, the BZT provokes a frequency warping phenomenon: after the BZT is performed, the digital frequency variable ω in the Z domain is not linearly but tangentially related to the analogue frequency variable ω' in the Laplace domain. If follows that the lower half of the digital frequency range [0:Fs/4] is expanded and the higher half [Fs/4:Fs/2] is compressed.

This warping effect is inherent to the BZT but it can be compensated to ensure that the digital filter has the desired cut-off frequency ω_c: Since we know the analytical relation between analogue and digital frequencies, we can calculate a pre-warped analogue frequency ω_c' whose digital equivalent after BZT is exactly the desired cut off frequency ω_c. This analogue cut off frequency ω_c' is used in the analogue expression of the filter in replacement of ω_c. Of course, this does not correct the non-linearity of the relation between analogue and digital frequencies but it ensures that digital frequencies are re-scaled so that the digital filter behaves according to its analogue counterpart around the desired cut off frequency. The prewarping is carried out by replacing ω_c in the analogue transfer function with ω_c'. The expression for ω_c' is given below:

$$\omega'_c = \tan\left(\frac{\omega_c}{2F_s}\right) \qquad \text{(Eq. 53)}$$

where ω_c is the desired cut off frequency, ω'_c is the prewarped analogue equivalent frequency and , Fs is the system sampling frequency. Once the BZT is applied to the prewarped analogue transfer function, the filter's coefficients vectors a and b are directly extracted from the expression for H(z). For simplicity of calculation, the frequency prewarping and the variable change of the BZT can even be combined in one and single expression given below:

$$s = \cot(\omega_c T/2) \cdot \frac{z-1}{z+1} \qquad \text{(Eq. 54)}$$

It is obvious that this combined transformation is easier and quicker to perform than the single pole simple element decomposition of the impulse invariant method. Additionally, this method has the simplicity of being entirely numerical. A disadvantage of the BZT however is that prewarping can only be applied for one single frequency. Consequently, the BZT should not be used to derive filters with multiple pass bands, like comb filters for example. This represents a limit to its

usefulness. Having said that, the BZT is the method of choice for earlydesign of simple IIR filters. For example, it is used in Matlab to produce the coefficients simulating a low pass Butterworth filter of order n:

$$[b,a] = butter(n,w); \quad \text{(Eq. 55)}$$

where b and a are the filter coefficients, n is the filter order and w is the cut off frequency. The butter function has a priori knowledge of the Butterworth transmittance and the cut off frequency w is fed into it.

5.14.3. Comparison of FIR and IIR filters and their applications

Roughly, FIR filters have the advantage of simplicity in design and implementation and the disadvantage of their need for large number of coefficients (hence longer convolution time) compared to IIR filters. For this reason, they can be delicate to use in real time applications. However, they are perfectly suited for off line work, a fortiori when zero phase filtering is required. FIR filters can have a linear phase variation, hence a constant phase delay whereas IIR filters have inherent non-linear phase variations, which can sometimes be linearly approximated in a small frequency range. IIR filters usually provide sharper cut off in their frequency responses, with the drawback of being destructive of the phase structure (i.e. the phase relations between harmonics of the input signal.

FIR filters can be used to design filters that have no physical equivalent, typically a moving average filter, an all band integrator, and an all band differentiator, or any other exotic prototype. In contrast, IIR filters are built and used in simulation of analogue filters.

5.14.3.1. Applications of FIR and IIR filters to the processing of FORP signals

Both IIR and FIR filters have been used to process FORP signals. Ideally, FIR filters should always be preferred for that such filter guarantee no distortion of the waveform due to dephased harmonics.

A zero phase filter based the bandpass Butterworth filter has been employed for removing the trend (one might prefer the term baseline excursion) and the noise of the cardiac plethysmographic signal. See section 12.4.1.1.

When real time filtering of the cardiac FORP signal was required, we found no other alternative but to implement an IIR bandpass filter (Bessel), which has an infinite impulse response. This is because, among IIR filters, Bessel filters offer minimal phase distortion and phase delay. Although

phase distortion could be observed in the filtered cardiac waveform, the overall phase delay of the signal was nearly constant, and minimal. See section 12.4.2.2.

FIR filters were used for respiratory monitoring with the FORP, and, thanks to the Wiener optimal filter method, the FIR vectors were defined to maximise the resemblance between the filtered FORP signal and the real respiratory signal. See section 11.3.

Chapter 6 MAGNETIC RESONANCE IMAGING

The principles of Magnetic Resonance Imaging (MRI) presented in this chapter permit one to understand the specific set of conditions imposed on any MRI compatible instruments by the Magnetic Resonance (MR) scanner.

Our interest in MRI in this research program was to assess the potential usefulness of the FORP as an MRI compatible cardiac heart rate measurement. The FORP seems to present itself naturally as a suitable technique to assist cardiovascular MRI because it is a non-metallic, non-invasive cardio-respiratory monitoring system. The development of this idea is the subject of chapter 12.

6.1. Introduction

MRI is a non-invasive, non-ionising imaging technique used primarily in medical settings to produce high quality images of the inside of the human body. It uses static and gradient magnetic fields to induce resonance in hydrogen nuclei. The study of signals emitted by the magnetised hydrogen nucleus allows the construction of an image. In comparison with X-ray Computed Tomography (CT), MRI is characterised by a high soft tissue contrast [Crooks et al. (1982)]. MRI is based on a spectroscopic technique called Nuclear Magnetic Resonance (NMR) operating in the Radio Frequency (RF) range of the electromagnetic spectrum.

6.2. Historical development

Felix Bloch and Edward Purcell, both of whom were awarded the Nobel Prize in 1952, discovered the magnetic resonance phenomenon independently in 1946. In the period between 1950 and 1970, Nuclear Magnetic Resonance (NMR) was developed and used for spectroscopic analysis. In 1971 Raymond Damadian showed that the nuclear magnetic relaxation times of tissues and tumours differed, thus motivating scientists to consider magnetic resonance for the detection of disease. Magnetic resonance imaging was first demonstrated on test tube samples that same year by Paul Lauterbur. In 1975 Richard Ernst proposed MRI using phase and frequency encoding, and the Fourier Transform. This technique is the basis of current MRI techniques. A few years later, in 1977, Raymond Damadian demonstrated MRI of the whole body. In this same year, Peter Mansfield developed the echo-planar imaging (EPI) technique. This technique was developed in later years to produce images at video rates (30 ms / image). Edelstein and co-workers demonstrated imaging of the body using Ernst's technique in 1980. A single image could be acquired in approximately five minutes by this technique. By 1986, the imaging time was reduced to about five seconds, without sacrificing too much image quality. In 1987 echo-planar imaging was used to perform real-time movie imaging of a single cardiac cycle. In this same year Charles Dumoulin was perfecting magnetic resonance angiography (MRA), which allowed imaging of flowing blood without the use of contrast agents. In 1991, Richard Ernst was rewarded for his achievements in pulsed Fourier Transform NMR and MRI with the Nobel Prize in Chemistry.

In 1993 functional MRI (fMRI) was developed. This technique allows the mapping of the function of the various regions of the human brain at work. The development of fMRI opened up a new application for EPI in mapping the regions of the brain responsible for thought and motor control.

6.3. Principles of magnetic resonance

6.3.1. Protons spin excitation

The human body is primarily fat and water, and contains approximately 63% of hydrogen atoms by proportion with other elements. Magnetic resonance imaging primarily produces the NMR signal from the hydrogen nuclei. The hydrogen nucleus is a proton. The proton has both an electric charge and a net spin angular momentum. In other words, it is a rotating charge and, consequently, it produces its own magnetic field and therefore behaves as a magnet.

When placed in an external magnetic field, the proton magnetic moment aligns itself with the direction of the external field. This is described as equilibrium magnetisation.

Considering a large ensemble of protons, a very small minority will align their spin anti-parallel with the direction of the external magnetic field, therefore in a position requiring a specific energy state that is higher than the energy state of those aligned parallel. All protons, either aligned parallel or anti–parallel with the external field precess at a frequency given by the Larmor equation:

$$f_L = \frac{\gamma B}{2\pi} \qquad \text{(Eq. 1)}$$

where f_L is the precession frequency, γ is the gyromagnetic ratio and B is the strength of the magnetic field surrounding the protons. The Larmor frequency is linearly related to the external magnetic field strength and variation in the field strength will cause protons to precess accordingly. The precession is comparable to a rotational movement of the proton's magnetic vector around the direction of the external magnetic field [Schild (1993)]. The gyromagnetic ratio value is 42.5 MHz/T for the hydrogen nucleus. External magnetic field strengths of 1.5 T to 2 T are common nowadays in MRI scanners[1] and consequently f_L is in the range 10 to 20 MHz, which corresponds to the Radio Frequency (RF) domain of the electromagnetic spectrum.

Let's consider a very small dot-shaped body tissue sample containing a few million protons. The vectorial sum of the magnetic vectors of each individual proton forms the so-called net magnetisation vector. Because the protons occur in very large numbers, their individual magnetic vectors are evenly distributed around the circular trajectory and subsequently their vectorial sum is

aligned with the external field. Because the quantity of anti-parallel protons is negligible at equilibrium compared to the number of parallel protons, the net magnetisation vector is aligned parallel with the direction of the external field applied on the tissue sample. This situation is represented below in a simplified manner.

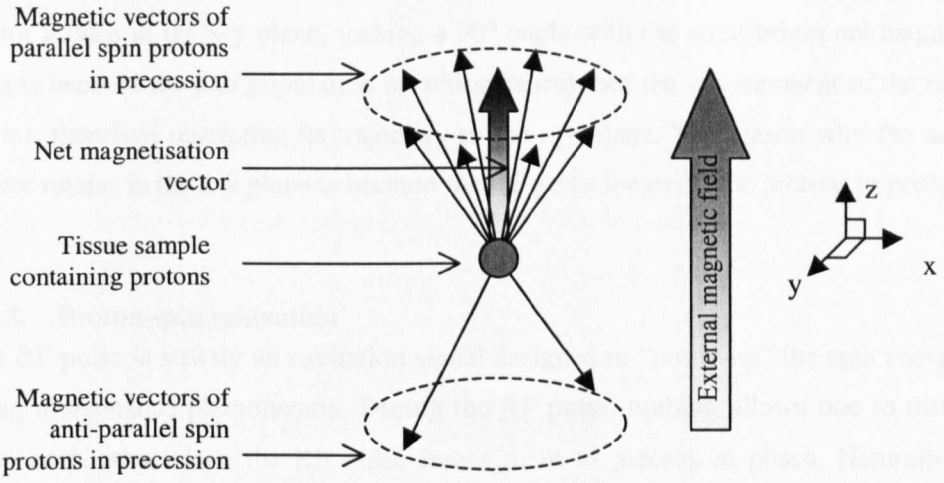

Figure 1:Simplified proton configuration at equilibrium magnetisation. The circles symbolise the trajectory of the proton magnetic vectors in precession.

The spin inversion energy is the energy required to provoke the transition of a proton spin from a low energy state (parallel alignment with external magnetic field) to a high energy state (anti-parallel alignment with external magnetic field). This energy is related to the Larmor frequency in the following manner:

$$\Delta E = h \cdot f_L \qquad \text{(Eq. 2)}$$

where ΔE is the energy difference between the parallel and anti-parallel spin, h is Planck's constant and f_L is the Larmor frequency. This equation tells us that, by sending a pulsed electromagnetic wave of frequency matching the Larmor frequency (commonly designated in MRI jargon as an RF pulse), we can provide the exact amount of energy required to invert the spin alignment of protons, turning them into anti-parallel spin protons. Therefore, by carefully controlling the time of exposure of the tissue sample to the RF pulse, the balance between low and high energy states in the proton spin population can be accurately controlled. The RF pulse also has the effect of making all exposed protons precess exactly with the same frequency and phase. The desired effect of altering the distribution of the proton spin population is to change the orientation of the net magnetisation vector.

---

1 Experimental MRI uses field strengths up to 10T, which allows high image resolution.

Two particular RF pulses are used in MRI.

The 180° pulse is simply an RF of duration sufficient to produce a complete spin population inversion, which results in the net magnetisation vector being aligned anti-parallel to the net magnetisation vector at equilibrium.

The 90° pulse has a smaller duration than the 180° pulse and is designed to provoke only half of a full spin population inversion. Immediately after exposure to this pulse, the net magnetisation vector rotates in the x-y plane, making a 90° angle with the equilibrium net magnetisation vector. This is because the half population inversion cancels out the z component of the net magnetisation vector, therefore restricting its trajectory to the x-y plane. The reason why the net magnetisation vector rotates in the x-y plane is because the RF pulse forces all the protons to precess in phase.

6.3.2. Proton spin relaxation

The RF pulse is strictly an excitation signal designed to "pump-up" the spin energy of protons by using a resonance phenomenon. During the RF pulse, nothing allows one to distinguish protons from each other since the RF pulse forces them to precess in phase. Naturally, what happens immediately after the RF pulse is that the proton spin population gradually returns to its original energy distribution, by re-emitting RF signals carrying the excess of spin energy.

To avoid the confusion between the RF pulse applied to the tissue sample to excite the proton spin population and the RF signal emitted by the same population during relaxation, the relaxation RF signals will now be called the MR signals. This makes sense because the MR scanner is designed to record and interpret these signals.

The characteristics of the MR signals emitted by a tissue sample are dependent on the trajectory of the net magnetisation vector during the protons "relaxation" period. For example, the trajectory of a net magnetisation vector recovering from a 90° RF pulse is a three dimensional dome-shaped spiral whose radius progressively decreases as the vector re-aligns itself with the external magnetic field. This is also known as the Free Induction Decay.

The trajectory of the net magnetisation vector during relaxation is influenced by two independent phenomena, which are now explained.

6.3.2.1. Spin-spin relaxation

Spin-spin relaxation relates to the progressive loss of phase coherence of the precessing protons: the proton population in the tissue sample gradually regains its original distribution of precession frequencies. This contributes to decrease the magnitude of the projection of the net magnetisation vector in the x-y (transverse) plane. For a 90° RF pulse, this projection of the FID is a spiral of decaying magnitude. The decay of magnitude in time typically follows an exponential model:

$$M_{xy} = M_{xy0} \cdot e^{-t/T_2} \qquad \text{(Eq. 3)}$$

where M_{xy} is the magnitude of the projection in the x-y plane, M_{xy0} is the initial magnitude of the projection (just before the RF pulse is switched off), t is the time and T_2 is the so-called spin-lattice relaxation time constant.

By fitting an exponential envelope to consecutive measurements of M_{xy}, the values of M_{xy0} and T_2 are identified. M_{xy0} is a measure of the density of protons and T_2 is a measure of how quickly the protons' phase coherence is lost during relaxation. Essentially, loss of phase coherence originates from inhomogeneities of the magnetic field in the surroundings (i.e. the lattice), causing the protons to precess with slightly different frequencies.

These inhomogeneities are characteristic of each particular type of body tissue. In a pure substance like de-ionised water for example, there are very few inhomogeneities and the magnetic field remains isotropic, resulting in a long T_2, in the range of 150 ms. On the other hand, water containing impurities will result in magnetic field anisotropy, hence a shorter T_2 in the range of 30 ms.

To sum up, the T_2 parameter is a measure of homogeneity in the spatial distribution of protons in a tissue sample.

6.3.2.2. Spin-lattice relaxation

Spin-lattice relaxation relates to the balance in the number of anti-parallel spin protons and parallel spin protons. During the relaxation period, the proton population is losing energy and therefore the number of anti-parallel protons is constantly decreasing. As the number of parallel spin protons increases, the envelope of the projection of the net magnetisation vector on the z-axis is also increasing (in the direction of the external magnetic field).

This growth reaches a plateau when the proton population has come back to its equilibrium magnetisation state. After a 90° RF pulse, the envelope of this second projection also follows an exponential model given by:

$$M_z = M_0(1 - e^{-t/T_1}) \qquad \text{(Eq. 4)}$$

where M_z is the envelope of the z-axis (longitudinal) projection of the net magnetisation vector, M_0 is the final magnitude of the net magnetisation vector at equilibrium, t is the time and T_1 is the so-called spin-lattice relaxation time constant.

Again, by fitting an exponential envelope to the consecutive measurements of M_z, the parameters M_0 and T_1 can be determined. M_0 is a measure of the density of protons in the sample tissue and T_1 relates to how quickly the excess of spin energy is returned to the surroundings.

In a substance containing a high density of mobile protons, such as water and fat, the MR signals emitted by anti-parallel spin protons are almost immediately re-absorbed by the parallel spin protons, turning them into anti-parallel spin protons and so on: as a result, M_z decays slowly and the value of T_1 is in the range of 2000 ms. Conversely, in a lower proton density tissue such as fat, the MR signals emitted by anti-parallel spin protons are rapidly absorbed by the lattice and the value of T_1 is much shorter (about 300 ms).

6.3.3. Image construction

The MRI scanner is equipped with a multitude of antenna coils disposed all around the patient that can perceive both longitudinal and transverse projections of the relaxation MR signals. The principle of MRI is to record and analyse these MR signals in order to spatially resolve the local variations in M_0, T_1 and T_2. The MR image is a map of the proton density M_0, weighted by the local variations in T_1 or T_2. Different types of image weighting allow one to distinguish between types of different body tissues.

6.3.3.1. Spatial resolution of body tissues

To spatially resolve the source tissues of all MR signals, the MR scanner first decomposes a volume into slices[2]. To encode the position of each slice along the z-axis, we use a gradient magnetic field[3] in the direction of the z-axis. The total magnetic field is then the sum of the permanent field and the gradient field.

Since the precession frequency is proportional to the total magnetic field intensity, the application of the gradient field results in each slice having a specific range of precession frequencies. This particular gradient field is generally termed slice encoding gradient field.

If the frequency of an MR signal falls into the frequency range of a particular slice, then the source of this MR signal is localised somewhere in this particular slice.

To locate the source of this MR signal within a slice, the MR scanner encodes x positions (columns) by using a second frequency encoding gradient field. This second gradient field provides subdivided precession frequencies varying within the limit of the slice frequency range. Again, the source column can be localised according to the frequency of its MR signal.

Finally, to locate the source of this MR signal within a column, the scanner does not use a third frequency encoding gradient field but a phase encoding gradient field. This technical exception is due to the insufficient spacing of the three precession frequency ranges in the x, y and z directions: a third frequency encoding gradient field would not permit the localisation of MR signals with sufficient precision and accuracy.

A solution to this problem is to increase the strength of the static magnetic field. However, a more powerful magnet is more expensive. The solution of using a phase encoding gradient field is ideal because it does not require its own frequency range to resolve the y-positions (rows). Instead, it operates within the frequency range of columns.

The principle of the phase encoding gradient field is to permit identification of the position of a source within a row according to the initial delay of the MR signal.

6.3.3.2. Pulse sequences

The idea behind pulse sequencing is to produce MR signals that can easily reveal the variations in M_{xy} and M_z of relaxing tissues. The pulse sequence refers to the strategic sequential combination of specific RF pulses, variations of the external magnetic field (gradient fields) and recording of MR signals. The nature and timing of the events of a pulse sequence are of paramount importance in the type of image obtained with the MR scanner.

---

2 The XY plane at a specific value of Z is termed a slice.

3 The gradient field is produced by specific magnetic coils.

Because of the weakness of the MR signal amplitude, it is necessary to repeat a pulse sequence many times in order to obtain an image of good quality: confidence in the image is improved by averaging the data acquired during repeated pulse sequences. Good image quality often requires long acquisition time.

Despite the efforts of subjects to remain still, long acquisition durations can induce a certain degree of blurring in the resulting image. This is especially true for body regions having autonomous activity like heart contractions for example. To increase image quality, we must increase the scanner sampling frequency. This is currently achieved with two approaches: 1) in multi-slice imaging, we run multiple pulse sequences simultaneously on different slices. 2) Fast pulse sequences are pulse sequences of shorter duration, hence producing an increase in the scanner sampling frequency. The principle of fast pulse sequencing is to use reduce the body tissue relaxation time by generating RF pulses that will not provoke a complete transverse magnetisation on the net magnetisation vector but a partial transverse magnetisation. This is referred to as "tilting" the net magnetisation vector.

6.3.3.3. MR signals analysis

The MR image is constructed by calculating the Inverse Fourier Transform of the MR signals emitted during relaxation: since the position of each MR signal within a slice is encoded with a phase and a frequency, the IFT lets us know where signals originate from by measuring their specific phases and frequencies.

6.4. Cardiac and cardiovascular imaging with the MR scanner.

Cardiovascular MRI permits the detection of a variety of heart diseases such as constrictive pericarditis, cardiac and para-cardiac masses, cardiomyopathy, valvular disease and ischemic heart disease. This is possible thanks to an arsenal of signal acquisition protocols and image construction methods. All these techniques are based on scanner synchronisation with the cardiac cycle of the patient.

6.4.1. Scanner trigger methods for Cardiovascular imaging

While imaging the respiratory system, it is possible to avoid motion blur by asking the patient to hold his/her breath for the duration of the scan. This procedure is limited by the patient's ability to maintain a stable chest position for up to 30 seconds but this does not constitute a complete impossibility. Because the heart cannot be stopped, MRI examination of the cardiovascular system is confronted by the inherent problem of heart movements and blood flow. In the present situation, cardiac MRI is made possible by means of triggering the scanner in synchrony with the R wave of the ECG signal.

Picking up an ECG signal in MRI conditions is made possible by using non-ferromagnetic materials for the electrodes and the leads. The ECG cable is shielded (except the extremities where the electrodes are connected) and has very high impedance in order to avoid burning the patient by stray induced currents[4]. Such currents are generated by the gradient magnetic fields. This is a very sensitive aspect of cardio-vascular MRI because an additional array of electromagnetic coils is placed on the patient's torso in order to produce stronger localised gradient magnetic fields. This obviously results in the production of a clearer image of the heart but also increases the probability of the patient receiving electrical burns. To limit such risks and get a clear ECG signal, the ECG electrodes are placed on the back of the patient. This makes the exact positioning of the electrodes a crucial part of the patient's preparation for the MRI examination.

In MRI conditions, the ECG signal contains artefacts due to the gradient magnetic fields and the movement of the magnetised blood. Intense denoising is required to precisely detect the R wave, hence to trigger the scanner.

The MRI ECG signal is nowadays the method of choice for cardiac gating but other methods exist as well. One of them is the Peripheral Pulse Gating (PPG) method. PPG measures blood flow velocity by means of optical detection through the fingertip. Other methods are the sphygmomanometer plethysmograph, which detects changes in limb distension with arterial pulsation, and laser-Doppler velocimetry which detects capillary blood volume of a superficial capillary bed (usually the tongue or ear lobe). All of these methods are not as accurate as ECG gating because they largely depend on the morphology of the patients cardiovascular systems.

6.4.2. Possible MR scanner trigger method based on the FORP

The FORP sensor is based on optical fibre and other non-metallic materials. This guarantees complete immunity to EM disturbances and consequently better functioning.

It has already been demonstrated that the FORP can help subjects to recover recovering a stable chest position during long MRI investigation [Raza (1998)]. Of more relevance to this thesis however, is the proposition that the FORP could also be used for synchronising the MR scanner with the cardiac cycle of a subject undergoing cardiac or cardiovascular MRI. This concept is presented and more thoroughly discussed in chapter 12.

4 This constitutes one of the primary causes of MRI accidents along with injuries caused by movement of objects that have become magnetised.

Chapter 7 PLETHYSMOGRAPHIC RESPIRATORY MONITORING

This chapter is an extended review and discussion of the existing techniques of plethysmographic respiratory monitoring. The materials presented here cover the principle of respiratory plethysmography, the calibration methods and the different plethysmographic respiratory instruments, including the previous version of the FORP.

7.1. Principle of respiratory plethysmography

Plethysmography is a measurement method where the volume of an object is estimated by the measurement of its external dimensions, according to an appropriate geometrical model describing the volume-surface area relation for the object.

The concept of determination of lung volume by respiratory plethysmography relies on measuring thoracic and abdominal cross sectional dimensions. Such an operation relies on the assumption that the ventilation system model is constituted of two compartments, as we will explain in this section. We implicitly verify that, during respiration, the chest and abdominal wall displacements indicate lung volume variation. It is shown by a phenomenological study that change of antero-posterior diameters of thorax and abdomen during respiration both contribute independently to the total respiratory volume. It maybe said that thorax and abdomen have their own degree of freedom in their contribution to the total respiratory volume. [Konno & Mead (1967), Smith & Mead (1981)] In other words, the total respiratory volume is expressed as the sum of thoracic and abdominal respiratory volumes. This is denoted as:

$$\Delta V = \Delta V_{TH} + \Delta V_{AB} \qquad \text{(Eq. 1)}$$

where ΔV is the respiratory volume change, ΔV_{TH} and ΔV_{AB} are respectively the thoracic and abdominal contributions to the respiratory volume change.

7.2. The Konno and Mead model

The idea of Konno and Mead was to model the ventilation system as two connected compartments, namely the thoracic and abdominal compartments, whose volumes are those of the thoracic and abdominal contributions to the total respiratory volume. One representation of this model due to Augousti [Augousti (1997)], based on modelling the compartments as cylinders, is given below:

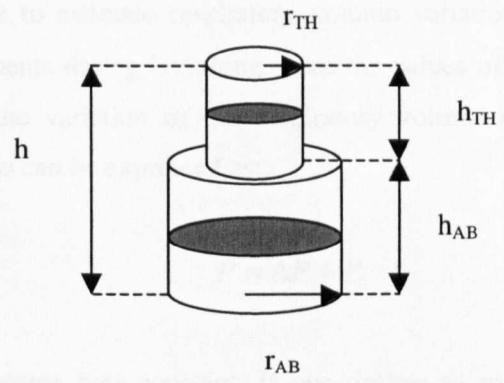

Figure 1: The connected compartment model. The greyed surfaces represent the thoracic and abdominal cross section surfaces, S_{TH} and S_{AB}. The initial Konno & Mead model did not imply any specific geometry. Augousti proposed that each compartment should be considered cylindrical.

In order to estimate ΔV, we must calculate ΔV_{TH} and ΔV_{AB} from their respective compartment dimensions. In Augousti's model, the height of each cylinder is initially constant. Therefore, the former two terms are further decomposed as follow:

$$\Delta V_{TH} = h_{TH} \cdot \Delta S_{TH} \quad \text{(Eq. 2)}$$

$$\Delta V_{AB} = h_{AB} \cdot \Delta S_{AB} \quad \text{(Eq. 3)}$$

where ΔS_{TH} and ΔS_{AB} represent thoracic and abdominal compartment cross sectional area variations respectively.

For convenience, we will assume in the following that the plethysmographic instrument has been carefully calibrated and returns a set of two signals called ΔP_{TH} and ΔP_{AB} that are measurements proportional to the cross sectional area variations ΔS_{TH} and ΔS_{AB}. However, since the height of each compartment is constant, ΔP_{TH} and ΔP_{AB} are also directly proportional to ΔV_{TH} and ΔV_{AB}. Hence, we can re-write the two equations above as:

$$\Delta V_{TH} = k_{TH} \cdot \Delta P_{TH} \quad \text{(Eq. 4)}$$

$$\Delta V_{AB} = k_{AB} \cdot \Delta P_{AB} \quad \text{(Eq. 5)}$$

where k_{TH} and k_{AB} are dimensionless proportionality constants serving as instrument gain calibration coefficients.

The last two equations and the Konno and Mead phenomenological equation can be combined and expressed as:

$$\Delta P = k_{TH} \cdot \Delta P_{TH} + k_{AB} \cdot \Delta P_{AB} \quad \text{(Eq. 6)}$$

where ΔP is the plethysmographic estimation of the relative ventilatory volume ΔV. This equation makes it possible to estimate respiratory volume variation from thoracic and abdominal cross section measurements during breathing, once the values of k_{TH} and k_{AB} are known. The term ΔP only represents the variation of the respiratory volume estimation P. The estimated absolute respiratory volume can be expressed as:

$$P = \Delta P + P_0 \quad \text{(Eq. 7)}$$

where P_0 is a volume bias constant. If one desires to match the plethysmographic respiratory volume estimation with V the respiratory volume measurements from a spirometer, it is necessary to determine P_0. By doing so, we can finally express the plethysmographic estimation of respiratory volume P, as:

$$P = k_{TH} \cdot \Delta P_{TH} + k_{AB} \cdot \Delta P_{AB} + P_0 \quad \text{(Eq. 8)}$$

This equation, once the constants have been adequately calibrated, allows us to generate a signal that is the plethysmographic estimation of the respiratory volume signal V, obtained from a spirometer, The addition of the constant term P_0 allows us advantageously to replace the terms ΔP_{TH} and ΔP_{AB} by their absolute counterparts P_{TH} and P_{AB}. The above equation is finally written as:

$$P = k_{TH} \cdot P_{TH} + k_{AB} \cdot P_{AB} + P_0 \quad \text{(Eq. 9)}$$

In this case, the values of k_{TH} and k_{AB} are necessarily different. The addition of the term P_0 does not constitute a change in the number of degrees of freedom because P_0 is constant during respiration. The expression above is sometimes also expressed as:

$$P = k_{AB} \cdot [M \cdot P_{TH} + P_{AB}] + P_0 \quad \text{(Eq. 10)}$$

where M is the ratio of the calibration constants, therefore expressing the contribution of the thoracic signal with respect to the abdominal signal.

7.3. Modified Konno and Mead model

In an attempt to provide physiological functionality to this phenomenological result, we propose an alternative interpretation of the relative independence of motion of thorax and abdomen during respiration. In the following, we do not consider the abdomen as a respiratory cavity, with the thorax only containing the whole respiratory volume. This approach has the merit of showing respect to torso anatomy and functionality while leading to the same final estimation of the respiratory volume. We will assume that the lungs are modelled as a compartment whose volume is expressed as:

$$V = h_{TH} \cdot S_{TH} \qquad \text{(Eq. 11)}$$

where V is the lung compartment volume (i.e. the respiratory volume), h_{TH} is the lung compartment height and S_{TH} is the lung compartment cross section area. The suffix TH is used to signify the fact that the lung compartment is positioned within the thorax. Exactly as in the previous model, such a volume equation implies that the corresponding compartment has a cross sectional area that will not vary along the axis of the compartment height. The study of the actions of the different respiratory muscles on the lungs shows that, during inspiration, the lungs can be expanded both along the vertical axis and in the horizontal plane by the independent actions of the diaphragm, the inter and intracostal muscles and the scalenes. This dimensional variation can be represented for the lung compartment by the independent variation of h_{TH} and S_{TH}. The independent variation of these two terms is equivalent to saying that the lung compartment is a ventilatory system with two degrees of freedom. Therefore, the respiratory volume variation can be expressed as:

$$\Delta V = h_{TH} \cdot \Delta S_{TH} + \Delta h_{TH} \cdot S_{TH} \qquad \text{(Eq. 12)}$$

where ΔV is the volume variation of the lung compartment, ΔS_{TH} is the cross sectional area change of the lung compartment and Δh_{TH} is the height change of the lung compartment.

The term $h_{TH} \Delta S_{TH}$ expresses the voluminal contribution to ΔV due to cross sectional area variation. Because lung cross sectional area variation ΔS_{TH} is more likely to occur as a result of costal muscle contractions rather than diaphragm displacement, $h_{TH} \Delta S_{TH}$ can be thought of as the equivalent of the term ΔV_{TH} in the Konno and Mead equation (see equation 1). Similarly, the term $\Delta h_{TH} S_{TH}$ is the voluminal contribution to ΔV due to height variation of the lung compartment and can be thought of as being the equivalent of ΔV_{AB} in the Konno and Mead equation because variation in lung height is more likely to be produced by diaphragm displacement rather than costal muscle contractions.

The abdomen is symbolised in the modified Konno & Mead model by adding a second compartment below the lung compartment. The volume of the abdominal compartment is expressed below as:

$$V_{AB} = h_{AB} \cdot S_{AB} \quad \text{(Eq. 13)}$$

where V_{AB} is the abdominal compartment volume, h_{AB} is the abdominal compartment height and S_{AB} is the abdominal compartment cross sectional area. The diaphragm effectively separates the compartments and can travel up and down freely, while ensuring that both compartments remain attached to it. The diaphragm position is identified by h_{AB}. The total height of both compartments is expressed as:

$$h = h_{TH} + h_{AB} \quad \text{(Eq. 14)}$$

where h is the total height. By making the assumption that h does not vary, we obtain:

$$\Delta h_{TH} = -\Delta h_{AB} \quad \text{(Eq. 15)}$$

Since one may assume that the abdomen has a constant volume[1], its volume variation is zero and therefore it follows that, for small variations of Δh_{AB} and ΔS_{AB}:

$$\frac{\Delta h_{AB}}{h_{AB}} = -\frac{\Delta S_{AB}}{S_{AB}} \quad \text{(Eq. 16)}$$

where Δh_{AB} is the height variation of the abdominal compartment and ΔS_{AB} is the cross sectional area variation of the abdominal compartment.

This signifies that diaphragm position variation results in cross sectional area variation of the abdominal compartment. By combining equations 15 and 16, we obtain an expression for the lung compartment height variation that relates to the cross sectional area of the abdominal compartment in the following way:

$$\Delta h_{TH} = \frac{h_{AB}}{S_{AB}} \cdot \Delta S_{AB} \quad \text{(Eq. 17)}$$

By substituting this above expression in equation 12, we finally obtain:

$$\Delta V = h_{TH} \cdot \Delta S_{TH} + \frac{S_{TH} \cdot h_{AB}}{S_{AB}} \cdot \Delta S_{AB} \quad \text{(Eq. 18)}$$

1 If we ignore the various effects of digestion, the abdomen has a constant mass and constant volumic mass, therefore a constant volume.

During normal quiet breathing, the cross section and height of each compartment only vary by small amounts and this permits us to consider non difference terms (h_{TH}, h_{AB}, S_{TH}, S_{AB}) in the equation above as constants. Hence we may write:

$$k_{TH} = h_{TH} \qquad \text{(Eq. 19)}$$

$$k_{AB} = \frac{S_{TH} \cdot h_{AB}}{S_{AB}} \qquad \text{(Eq. 20)}$$

It follows that the final expression from our model above now exactly matches the final expression from the model developed by Konno and Mead. This amendment to the two compartment approach to ventilation system modelling has the advantage of providing a physical basis for the physiological phenomenon of independent chest and abdominal antero-posterior diameter variations observed by Konno and Mead during respiration.

7.4. Modified Konno & Mead model with linear dependence of thoracic height variation on thoracic cross sectional area variation.

Although the thoracic compartment height h_{TH} and its cross sectional area S_{TH} have been treated as mathematically independent so far, physiological considerations suggest that they become linked for large respiratory volume variations: the scalenes are respiratory muscles that lift the ribcage near the end of a large inhalation. [Pocock and Richards (1999)] This is no loss of generality: it will be shown that thoracic height variations provoked by abdominal height variation or thoracic cross sectional area variation are two independent phenomena. Hence, a change in thoracic cross sectional area does not provoke a variation in abdominal height and the model still has two degrees of freedom. It was first proposed by Augousti [Augousti (1997)] to establish a linear dependence of the thoracic height variation Δh_{TH*} (induced by the scalenes) on the cross sectional area variation ΔS_{TH}. This is expressed as follows:

$$\Delta h_{TH*} = \lambda \cdot \Delta S_{TH} \qquad \text{(Eq. 21)}$$

where λ is a proportionality constant. In respect of the physiological functionality of the torso, we will apply the above expression to the modified Konno and Mead model previously described in section 7.3. (However, similar results would be obtained with the original model.)

Because Δh_{TH*} is independent on the compartment height h_{TH}, the total height of both compartments is now written as:

$$h = h_{TH} + \Delta h_{TH*} + h_{AB} \qquad \text{(Eq. 22)}$$

The following expression is still valid because the thoracic height variation Δh_{TH*} induced by the scalenes is independent of the thoracic height variation induced by the diaphragm.

$$\Delta(h_{TH} + h_{AB}) = 0 \qquad \text{(Eq. 23)}$$

hence

$$\Delta h_{TH} = -\Delta h_{AB} \qquad \text{(Eq. 24)}$$

The new thoracic height variation $\Delta h_{TH'}$ is the combination of the diaphragm induced thoracic height variation Δh_{TH} and the scalene induced thoracic height variation Δh_{TH*}. This is expressed as:

$$\Delta h_{TH'} = \Delta h_{TH*} + \Delta h_{TH} \qquad \text{(Eq. 25)}$$

hence

$$\Delta h_{TH}' = \Delta h_{TH*} - \Delta h_{AB} \qquad \text{(Eq. 26)}$$

and

$$\Delta h_{TH'} = \lambda \cdot \Delta S_{TH} - \Delta h_{AB} \qquad \text{(Eq. 27)}$$

This new expression of the thoracic compartment height variation replaces the previous one to calculate the respiratory volume variation. The respiratory volume variation is now expressed as:

$$\Delta V = h_{TH} \cdot \Delta S_{TH} + \Delta h_{TH'} \cdot S_{TH} \qquad \text{(Eq. 28)}$$

and it follows by combining equation 16 and 28 that:

$$\Delta V = [h_{TH} + \lambda \cdot S_{TH}] \cdot \Delta S_{TH} + [\frac{S_{TH} \cdot h_{AB}}{S_{AB}}] \cdot \Delta S_{AB} \qquad \text{(Eq. 29)}$$

All the terms in between the square brackets are constants implying that, after simplification, the above expression for the respiratory volume variation is equivalent to the one given by Konno and Mead. Therefore, the addition of linear dependence of the thoracic height on the thoracic cross sectional area does not modify the generality of the model. We conclude that, if such a dependence exists in reality, it would not create additional discrepancies because the model naturally takes it into account.

The approach initially suggested by Augousti was to determine the effect of a linear dependence of height on cross sectional area at both thoracic and abdominal levels and our present approach differs from this. However, given that the dependence that was introduced is invisible in the final

equation of the model, we can see retrospectively that our approach does not constitute a loss of generality compared to Augousti's since they both lead to the same result.

7.5. Modified Konno & Mead model with quadratic dependence of thoracic height variation on thoracic cross sectional area variation

The model presented in the previous section has the anatomical interest of establishing a dependence between the thoracic height and thoracic cross sectional area and the nature of this dependence was assumed to be linear. Augousti later proposed using a quadratic dependence of the height on cross sectional area for both thoracic and abdominal compartments. The assumption here is that the shape of a parabolic curve should ensure that the dependence of height on cross sectional area remains minimal for small cross section variations and becomes apparent for large cross section values. Further physiological considerations of the mode of action of the scalenes suggests that their assistance during inspiration becomes more pronounced near the end of the inhalation period and we can reason that a quadratic dependence of height on cross sectional area might be helpful in simulating the actions of the scalenes. The existence of such a dependence for the abdominal compartment is a matter of discussion as nothing indicates that this dependence might help in simulating any particular muscular action related to abdominal height variations. As far as the Modified Konno and Mead model abdomen is concerned, this quadratic dependence cannot be introduced because the height and cross sectional area variations are intrinsically linearly related. In terms of our modified Konno and Mead model, the quadratic dependence at thoracic level is denoted as:

$$\Delta h_{TH*} = \alpha \cdot \Delta S_{TH}^2 \qquad \text{(Eq. 30)}$$

Hence, the total thoracic height variation is now:

$$\Delta h_{TH'} = \alpha \cdot \Delta S_{TH}^2 - \Delta h_{AB} \qquad \text{(Eq. 31)}$$

And the model equation is now expressed as:

$$\Delta V = [\alpha \cdot S_{TH}] \cdot \Delta S_{TH}^2 + [h_{TH}] \cdot \Delta S_{TH} + [\frac{S_{TH} \cdot h_{AB}}{S_{AB}}] \cdot \Delta S_{AB} \qquad \text{(Eq. 32)}$$

We will note that the terms within the square brackets are constants that can be generalised to include the gain correction of each plethysmographic channel during calibration and thus may be re-written as:

$$P = l_{TH} \cdot P_{TH}^2 + k_{TH} \cdot P_{TH} + k_{AB} \cdot P_{AB} + P_0 \qquad \text{(Eq. 33)}$$

where l_{TH} is the calibration constant of the quadratic term with unit m^{-3}, and k_{TH} and k_{AB} are the dimensionless calibration constants of the linear terms. Augousti's initial suggestion to introduce a quadratic dependence at both thoracic and abdominal levels led to the following model equation:

$$P = l_{TH} \cdot P_{TH}^2 + k_{TH} \cdot P_{TH} + l_{AB} \cdot P_{AB}^2 + k_{AB} \cdot P_{AB} + P_0 \qquad \text{(Eq. 34)}$$

We can easily see that both model equations can be made equivalent to each other by forcing l_{AB} to a zero value.

In this section, we have seen that, to simulate the mode of action of the scalenes, the order of the dependence of height on cross sectional area was chosen as an integer superior to 1. The second order seems appropriate because it leads to simple model equation, while producing the same type of effect as higher orders. However, there is no anatomical reason that justifies the use of order 2 and we are entitled to wonder whether orders higher than 2 could yield better calibration results. Unfortunately, this quest for better calibration results is hindered by a signal processing problem: increasing power orders generate plethysmographic signals with rapidly increasing unpredictability due to the high amplification gain of such orders: The noise components of P_{th} and P_{ab} become too important..

7.6. Calibration of respiratory plethysmographic instruments

Calibration is the process of determining adequate values for the k_{TH} , k_{AB} and V_0 constants of the respiratory model equation. There are many reported methods of calibrating respiratory plethysmograph instruments, such as the RIP (see section 7.7.1) and the FORP, against direct measurements of respiratory volume [Sackner et al (1989), Loveridge et al (1983), Stagg et al (1978), Mead et al (1967)] and equally as many reviews for validation of these methods [Chada et al (1982), Stroberg et al (1993), Hudgel et al (1984a), Revow et al (1987)]. All these methods have evolved from the assumption first pointed out by Konno and Mead, that the respiratory system can be thought of as consisting of two independently moving compartments; the thorax and abdomen, and that the volume within each compartment is linearly related to the cross sectional area of that compartment [Konno & Mead (1967)].

If the relative contribution of each of these two compartments to the tidal volume remained constant during respiration, then predictions of tidal volume would require only the calibration of one of the compartments (to respired volume). In practice however, the contribution of thorax and abdomen may change [Zimmerman et al (1982)] and therefore it is necessary to compute the relationship of volume to cross sectional area for each of these independent compartments and then sum the two contributions.

It is important to note that due to the variability in the relative contributions from the thorax and the abdomen, it is necessary to determine the relationship between the two over the whole range of the

lung volume. In any case, coefficients retrieved from the calibrations are only valid for the posture in which they were performed. Separate calibrations are necessary for standing, sitting, and supine postures. A review of the calibration methods utilised is outlined below.

7.6.1. Calibration methods: Linear Isovolume

During breathing, the ventilation system model is said to have two degrees of freedom as thoracic and abdominal walls move freely without interdependence. When the ventilation system is closed, which occurs when the mouth and nose are prevented from exchange of air with the surroundings, then air may only be shifted from one compartment to the other and reciprocally, establishing a dependence in the volume of each compartment. This is the situation of isovolume manoeuvre, in which the respiratory volume variation is zero. In the terms of the Konno and Mead model, this is expressed as:

$$\Delta P = k_{TH} \cdot \Delta P_{TH} + k_{AB} \cdot \Delta P_{AB} = 0 \qquad \text{(Eq. 35)}$$

hence:

$$M \cdot \Delta P_{TH} + \Delta P_{AB} = 0 \qquad \text{(Eq. 36)}$$

where M is the ratio of the thoracic and abdominal calibration constants and, finally:

$$M = -\frac{\Delta P_{AB}}{\Delta P_{TH}} \qquad \text{(Eq. 37)}$$

This signifies that, during an isovolume manoeuvre, the ratio of the constants k_{TH} and k_{AB} is known by calculating the ratio of abdominal and thoracic plethysmographic signal variations. More precisely, the expression specifies that we can obtain a new value for the ratio of calibration constants for each new data pair ($\Delta P_{AB,n}$; $\Delta P_{TH,n}$). A statistically meaningful value of this ratio can be obtained by averaging over all of its measured values. This is expressed as:

$$M_{MEAN} = -\sum_{n=0}^{N-1} \frac{\Delta P_{AB,n}}{\Delta P_{TH,n}} \qquad \text{(Eq. 38)}$$

If one is concerned simply with obtaining a signal that reproduces an output that is proportional to respiratory volumes variation, then the isovolume manoeuvre is sufficient in itself. The following expression relates how the result of isovolume calibration is employed during normal breathing to generate a signal that constitutes an estimation proportional to the relative respiratory volume:

$$M_{MEAN} \cdot \Delta P_{TH} + \Delta P_{AB} \propto \Delta V \qquad \text{(Eq. 39)}$$

This method of relative calibration has the advantage of being completely non-invasive. However, a second calibration is required against a system measuring the absolute volume exchanges at the mouth (such as a spirometer or a pneumotachometer) in order to calibrate the plethysmographic system for absolute ventilatory volume measurements. By dividing the relative respiratory volume signal obtained from a spirometer by its relative plethysmographic estimation obtained by isovolume calibration, this second calibration determinates the value of k_{AB} and consequently k_{TH} can also be calculated. Finally, the volume bias constant can be obtained if necessary by minimising the difference between the relative respiratory volume estimation ΔV_{est} and the respiratory volume measurement V.

It can be shown that the ratio of calibration constants k_{TH} and k_{AB} is dependent on the existing volume in each compartment during isovolume calibration and will therefore be expected to depend on posture.

Having calibrated in this fashion, one must then assume that the proportion of air in each compartment will remain the same during normal breathing, which is not always a fair assumption. [Augousti (1997)]. More importantly perhaps, the use of isovolume calibration is somehow restricted to trained, healthy and co-operative subjects as the isovolume manoeuvre requires paradoxical torso movements that are unnatural. There have been reported cases of subjects being unable to produce the isovolume manoeuvre [Mead et al (1967)].

7.6.2. Calibration methods: Quadratic Isovolume

The isovolume method can also be applied to model equations with quadratic dependence of the compartmental height on cross sectional area, either at thoracic level only or at both thoracic and abdominal levels. (See the quadratic models in section 7.4) The latter model equation will be used here to keep the largest scope possible for the problem of calibration. We will remember that, in Augousti's approach, we have:

$$\Delta P = l_{TH} \cdot \Delta P_{TH}^2 + k_{TH} \cdot \Delta P_{TH} + l_{AB} \cdot \Delta P_{AB}^2 + k_{AB} \cdot \Delta P_{AB} \qquad \text{(Eq. 40)}$$

In the case of isovolume ventilation, no air can circulate through mouth or nose and therefore the respiratory volume corresponds to the constant volume of air trapped in the thoracic and abdominal compartments. For this reason, the variation of the plethysmographic signal P is forced to zero.

$$l_{TH} \cdot \Delta P_{TH}^2 + k_{TH} \cdot \Delta P_{TH} + l_{AB} \cdot \Delta P_{AB}^2 + k_{AB} \cdot \Delta P_{AB} = 0 \qquad \text{(Eq. 41)}$$

which can be recast as:

$$\frac{k_{TH}}{l_{TH}} \cdot \Delta P_{TH} + \frac{l_{AB}}{l_{TH}} \cdot \Delta P_{AB}^2 + \frac{k_{AB}}{l_{TH}} \cdot \Delta P_{AB} = -\Delta P_{TH}^2 \qquad \text{(Eq. 42)}$$

or, more simply:

$$\alpha \cdot \Delta P_{TH} + \beta \cdot \Delta P_{AB}^2 + \gamma \cdot \Delta P_{AB} = -\Delta P_{TH}^2 \qquad \text{(Eq. 43)}$$

where α is k_{TH}/l_{TH}, β is l_{AB}/l_{TH}, and γ is k_{AB}/l_{TH}. Solving this single equation is impossible as we have three variables, namely α, β and γ. By providing two more equations, we are a in position to calibrate the plethysmographic system. Two such equations are obtained by recording two other interlaced sets of data during the same isovolume manoeuvre. This leads to a system of 3 equations (each identified with a suffix from one to three) with 3 variables which has the following matrix form expression:

$$\begin{bmatrix} \Delta P_{TH,1} & \Delta P_{AB,1}^2 & \Delta P_{AB,1} \\ \Delta P_{TH,2} & \Delta P_{AB,2}^2 & \Delta P_{AB,2} \\ \Delta P_{TH,3} & \Delta P_{AB,3}^2 & \Delta P_{AB,3} \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} -\Delta P_{TH,1}^2 \\ -\Delta P_{TH,2}^2 \\ -\Delta P_{TH,3}^2 \end{bmatrix} \qquad \text{(Eq. 44)}$$

In compact matrix form, the expression above is written:

$$S \cdot K = C \qquad \text{(Eq. 45)}$$

where K is the left hand side vector containing the calibration constants, S is a square matrix and C is the right hand side vector. Finally, a matrix inversion is required to obtain the calibration constants and we write:

$$K = S^{-1} \cdot C \qquad \text{(Eq. 46)}$$

Given that the inversion of S and its multiplication to C would be carried out numerically with a computer[2], no explicit literal expression is reproduced here for K. As in the case of the linear isovolume calibration method, to obtain statistically meaningful values of α, β and γ, we need to re-calculate the vector K for each new set of data ($\Delta P_{AB,n}$; $\Delta P_{TH,n}$; $\Delta P_{AB,n+1}$; $\Delta P_{TH,n+1}$; $\Delta P_{AB,n+2}$; $\Delta P_{TH,n+2}$) and average all instances of K. This is expressed as follows:

2 This calibration method had been previously implemented by A. Raza for real time operation, using Visual Basic.

$$K_{MEAN} = \sum_{n=0}^{(N-2)/3} K_{3n+1} \qquad \text{(Eq. 47)}$$

We can clearly see that this method rapidly becomes computationally intensive and can effectively only be implemented with the help of computational software. As for the linear isovolume method, we are able to determine all the calibration constants except two (namely l_{TH} and P_0), which makes this calibration method generate a plethysmographic signal with the correct wave shape but that is not to scale and in position with the corresponding spirometric signal.

To fully calibrate the plethysmographic system, a second calibration is necessary against a spirometer, whereby therefore l_{TH} and P_0 maybe determined.

7.6.3. Calibration methods: Multiple Linear Regressions

Multiple Linear Regression is a calibration method based on finding the least mean square difference between the respiratory volume signal V (as delivered by a spirometer) and the plethysmographic signal P [Chada et al. (1982)]. This method requires the use of a spirometer and is therefore invasive, as opposed to the isovolume method. However, it imposes no restrictions on the type of breathing used during the calibration. In particular, it can also be used with the isovolume manoeuvre, thereby providing a way of comparing the results of linear isovolume calibration against a more mathematically elaborate method. To differentiate the relative contribution of thoracic and abdominal compartment volumes to the spirometer volume measurement, subjects are instructed to breathe predominantly by moving the ribcage for a series of 10 breaths and then by moving the abdomen for a series of 10 breaths. The processing of the P_{TH}, P_{AB} and V to obtain the coefficients k_{TH} and k_{AB} then starts with the definition of the mean square deviation Q as:

$$Q = \sum_{n=0}^{N-1} [V_n - P_n]^2 \qquad \text{(Eq. 48)}$$

where P_n is a sequence of sampled values constituting the discrete plethysmographic estimation of the respiratory signal P, V_n is the sequence of sampled values corresponding to the spirometer signal V and N is the number of sampled values in each signal. P_n is defined as:

$$P_n = k_{TH} \cdot P_{TH,n} + k_{AB} \cdot P_{AB,n} + P_0 \qquad \text{(Eq. 49)}$$

The expansion of Q and its minimisation with respect to k_{TH}, k_{AB} and P_0 leads to the following system of simultaneous equations expressed in matrix form:

$$\begin{bmatrix} \sum P_{TH,n}^2 & \sum P_{TH,n} \cdot P_{AB},n & \sum P_{TH,n} \\ \sum P_{TH,n} \cdot P_{AB,n} & \sum P_{AB,n}^2 & \sum P_{AB,n} \\ \sum P_{TH,n} & \sum P_{AB,n} & N \end{bmatrix} \cdot \begin{bmatrix} k_{TH} \\ k_{AB} \\ P_0 \end{bmatrix} = \begin{bmatrix} \sum VP_{TH,n} \\ \sum VP_{AB,n} \\ \sum V_n \end{bmatrix} \quad \text{(Eq. 50)}$$

This expression can be re-written as:

$$S \cdot K = C \quad \text{(Eq. 51)}$$

where K is the left hand side vector containing the calibration constants, S is a square matrix and C is the right hand side vector. Finally, a matrix inversion is required to obtain the calibration constants and we write:

$$K = S^{-1} \cdot C \quad \text{(Eq. 52)}$$

Given that the inversion of S and its multiplication to C is carried out numerically in Matlab, no explicit literal expression is presented here for K. The sums used in the expressions for S and C concern the first N samples of the concerned variables. Depending on the value of N, this method can rapidly become computationally intensive and obviously necessitates a computer for the calculation of its solution. The use of the term P_0 has been added to the original version of this method [Sackner at al (1980)].

One drawback associated with this calibration method is that it utilises no a priori knowledge of the anatomical functionality of the chest and in particular of the equilibrium of thoracic and abdominal voluminal contribution to the respiratory volume. The method is simply required to fit three signals together with a view to obtaining a plethysmographic signal that resembles as closely as possible its spirometric equivalent. Calibration results obtained with this method are presented in section 11.2.2. There is no restriction on the range of possible values of the calibration constants and it can sometimes be noticed that the method will almost completely ignore one of the plethysmographic signals, judging that its contribution to the simulation of the spirometric signal is "unrequired". Even worse, it has been observed in some cases that the method was returning negative calibration values[3].

The ambiguity of this regressive calibration method is that a "good looking" estimation of the spirometer signal does not necessarily mean that the calibration constants have an ideal value. This is more thoroughly discussed in section 11.2.3.

---

3 A possible physiological sense for a negative calibration value is that the patient could be adopting an "almost isovolumic" breathing pattern. For example, if the respiration is predominantly thoracic, then the abdomen might be driven by the thorax, therefore generating a negative abdominal calibration constant. A more plausible explanation however is that the regressive calibration procedure "makes the choice" to use negative calibration values in an attempt to reshape one of the plethysmographic signals using the other.

7.6.4. Calibration methods: Multiple Quadratic Regression

This method is the quadratic version of the previous method (presented in section 7.5.3) and seeks to determine the best fit of the plethysmographic signals to the spirometric signal by adapting the calibration constants in the plethysmographic model equation. We will remember that two versions of the quadratic plethysmographic model equation were presented, one with both thoracic and abdominal quadratic dependence of height on cross sectional area (Augousti's equation) and another with only a thoracic quadratic dependence of height on cross sectional area. Augousti's equation has the largest scope and will therefore be used here. (The other model equation can be obtained from Augousti's by simply putting the calibration factor related to quadratic dependence at abdominal level to zero.)

Augousti's equation was expressed as:

$$P_n = l_{TH} \cdot P_{TH,n}^2 + k_{TH} \cdot P_{TH,n} + l_{AB} \cdot P_{AB,n}^2 + k_{AB} \cdot P_{AB,n} + P_0 \qquad \text{(Eq. 53)}$$

We define the mean square deviation Q as:

$$Q = \sum_{n=0}^{N-1} [V_n - P_n]^2 \qquad \text{(Eq. 54)}$$

where V_n is the spirometric signal and N is the total number of samples. The expantionand minimisation of Q with respect to l_{TH}, k_{TH}, l_{AB} and k_{AB} leads to the following system of five simultaneous equations expressed in expanded matrix form below:

$$\begin{bmatrix} \sum P_{TH,n}^2 & \sum P_{TH,n}^3 & \sum P_{TH,n} \cdot P_{AB,n} & \sum P_{TH,n} \cdot P_{AB,n}^2 & \sum P_{TH,n} \\ \sum P_{TH,n}^3 & \sum P_{TH,n}^4 & \sum P_{TH,n} \cdot P_{AB,n}^2 & \sum P_{TH,n}^2 \cdot P_{AB,n}^2 & \sum P_{TH,n}^2 \\ \sum P_{TH,n} \cdot P_{AB,n} & \sum P_{TH,n} \cdot P_{AB,n}^2 & \sum P_{AB,n}^2 & \sum P_{AB,n}^3 & \sum P_{AB,n} \\ \sum P_{TH,n} \cdot P_{AB,n}^2 & \sum P_{TH,n}^2 \cdot P_{AB,n}^2 & \sum P_{AB,n}^3 & \sum P_{AB,n}^4 & \sum P_{AB,n}^2 \\ \sum P_{TH,n} & \sum P_{TH,n}^2 & \sum P_{AB,n} & \sum P_{AB,n}^2 & N \end{bmatrix} \cdot \begin{bmatrix} k_{TH} \\ l_{TH} \\ k_{AB} \\ l_{AB} \\ P_0 \end{bmatrix} = \begin{bmatrix} \sum S_n \cdot P_{TH,n} \\ \sum S_n \cdot P_{TH,n}^2 \\ \sum S_n \cdot P_{AB,n} \\ \sum S_n \cdot P_{AB,n}^2 \\ \sum S_n \end{bmatrix}$$

(Eq. 55)

This system can be written in compact form as:

$$S \cdot K = C \qquad \text{(Eq. 56)}$$

where K is the left hand side vector containing the calibration constants, S is a 5 by 5 square matrix and C is the right hand side vector. Finally, the square matrix S is inverted, multiplied to C and the vector K is determined. The above matrix system can easily be modified to calibrate the model equation without quadratic dependence of abdominal height on cross sectional area. The required modifications concern the removal of the term P^2_{AB} and its associated constant l_{AB}, from the system

above, since neither of them exists in the current model. Therefore, we can re-write the system above after simplification as:

$$\begin{bmatrix} \sum P_{TH,n}^2 & \sum P_{TH,n}^3 & \sum P_{TH,n} \cdot P_{AB,n} & \sum P_{TH,n} \\ \sum P_{TH,n}^3 & \sum P_{TH,n}^4 & \sum P_{TH,n}^2 \cdot P_{AB,n} & \sum P_{TH,n}^2 \\ \sum P_{TH,n} \cdot P_{AB,n} & \sum P_{TH,n}^2 \cdot P_{AB,n} & \sum P_{TH,n}^2 \cdot P_{AB,n}^2 & \sum P_{AB,n} \\ \sum P_{TH,n} & \sum P_{TH,n}^2 & \sum P_{AB,n} & N \end{bmatrix} \cdot \begin{bmatrix} k_{TH} \\ l_{TH} \\ k_{AB} \\ P_0 \end{bmatrix} = \begin{bmatrix} \sum S_n \cdot P_{TH,n} \\ \sum S_n \cdot P_{TH,n}^2 \\ \sum S_n \cdot P_{AB,n} \\ \sum S_n \end{bmatrix}$$

(Eq. 57)

This system can be written in compact form as:

$$S \cdot K = C \qquad \text{(Eq. 58)}$$

where K is the left hand side vector containing the calibration constants, S is a 4 by 4 square matrix and C is the right hand side vector. The square matrix S is inverted, multiplied to C and the vector K is finally determined.

7.6.5. Calibration methods: Qualitative Diagnostic Calibration

The Qualitative Diagnostic Calibration (QDC) method was presented by Sackner [Sackner et al (1989)] in an attempt to avoid the need for a volumetric respiratory measurement instrument to calibrate the linear plethysmographic equation, but without requiring isovolume manoeuvres. The obvious advantage of this method is its complete non-invasiveness and absence of conscious co-operation from the subject: he /she is only required to breathe spontaneously and regularly as would happen after a period of relaxation or during sleep. Theoretically, if a subject could breath with a constant respiratory volume during normal breathing then the population of peak-to-peak, breath-by-breath respiratory volumes would have a constant single value. Consequently, the standard deviation of such a population would be zero. This initial assumption is the basis of the QDC method. This is expressed as:

$$\sigma(p2p(\Delta V)) = 0 \qquad \text{(Eq. 59)}$$

where p2p(ΔV) represents the population of breath by breath, peak-to-peak respiratory volumes, also known as the tidal volumes, and σ represents the standard deviation. We will recall that a linear respiratory plethysmographic instrument lets us estimate the respiratory volume ΔV in the following fashion:

$$\Delta P = k_{AB} \cdot (M \cdot \Delta P_{TH} + \Delta P_{AB}) \qquad \text{(Eq. 60)}$$

where ΔP is the plethysmographic estimation of ΔV. The tidal volume p2p(ΔV) is therefore estimated using the following expression:

$$p2p(\Delta P) = k_{AB} \cdot (M \cdot p2p(\Delta P_{TH}) + p2p(\Delta P_{AB})) \qquad \text{(Eq. 61)}$$

where p2p(ΔP_{TH}) and p2p(ΔP_{AB}) are the thoracic and abdominal peak-to-peak plethysmographic amplitudes. In this equation, the peak-to-peak function p2p has been distributed on the right hand side. This is possible given the condition that ΔP_{TH} and ΔP_{AB} have their maxima and minima at the same moment, so that ΔP_{TH} and ΔP_{AB} are in phase. In practice, this is not necessarily the case.

We have said that the standard deviation of the tidal volumes is assumed to be zero and consequently its plethysmographic analogue should behave accordingly. This is written as:

$$\sigma(p2p(\Delta P)) = 0 \qquad \text{(Eq. 62)}$$

and, hence:

$$\sigma(M \cdot p2p(\Delta P_{TH}) + p2p(\Delta P_{AB})) = 0 \qquad \text{(Eq. 63)}$$

It is interesting to note the structural similarity between the above equation and the linear isovolume equation described in section 7.5.1.

By treating ΔP_{TH} and ΔP_{AB} as statistically independent variables (which is, in fact, rather questionable) with a zero average, the variance of a sum becomes the sum of variances and Sackner was able to re-cast the above equation in the following manner:

$$M \cdot \sigma(p2p(\Delta P_{TH})) + \sigma(p2p(\Delta P_{AB})) = 0 \qquad \text{(Eq. 64)}$$

Strictly speaking, this equation must be wrong because, in the first place, the sum of the variance of two statistically independent variables cannot be zero unless both variances are zero themselves. This error is a consequence of the initial assumption and the previous assumption of variable independence. Sackner reasoned that the best approximation of a zero variance is a minimal variance. The variances of the populations p2p(ΔP_{TH}) and p2p(ΔP_{AB}) can indeed be minimised by removing outlying values from the data set. (The statistical exclusion criterion used for this purpose is explained later in section 7.5.5.1.) Finally, M can be determined as follows:

$$M = -\frac{\sigma(p2p(\Delta P_{AB}))}{\sigma(p2p(\Delta P_{TH}))} \qquad \text{(Eq. 65)}$$

An *a posteriori* analysis of the equation above is that, to form the estimation of the respiratory volume, the abdominal and thoracic signals are balanced by the ratio of their respective standard deviation. In other words, if a signal has a standard deviation larger than the other signal, the contribution of this first signal to the estimation of respiratory volume is diminished. Hence, this procedure ensures that the estimation of respiratory volume has the most constant amplitude range possible

.

7.6.5.1. Statistical exclusion method for the minimisation of the variances of p2p(ΔP_{TH}) and p2p(ΔP_{AB})

Because the QDC method is based on the assumption that the tidal volume is constant during natural breathing, an *estimation* of the tidal volume must also be constant in the same condition of breathing. The best estimation of the tidal volume is its most probable value, which in our case corresponds to its average[4].

Given the variable independence, and the fact that the value of M is still undetermined at this stage (the statistical exclusion is performed before the calculation of M), the tidal volume cannot be estimated from equation 60. However, a biased but accurate estimator of the tidal volume is the average over all breaths of p2p(ΔP_{TH}) and p2p(ΔP_{AB}). The estimator is therefore termed the uncalibrated estimation of the tidal volume.

The analysis of the deviation of the breath to breath sum of p2p(ΔP_{TH}) and p2p(ΔP_{AB}) from the uncalibrated tidal volume estimation leads to the identification of outlying data. By carefully defining a threshold (one standard deviation from the mean, typically), it is possible to decide whether or not data should be included in the calculation of M.

7.6.5.2. Algorithm of the QDC method

The algorithm for implementation of the QDC method is now described: following an acquisition session of duration approximately 5 minutes, the population of p2p(ΔP_{TH}) and p2p(ΔP_{AB}) is formed by deducing the peak to peak amplitudes of the plethysmographic signals for each breath. Next is the calculation of the average sum of p2p(ΔP_{TH}) and p2p(ΔP_{AB}) and the exclusion of outlying data points. Finally, the remaining populations p2p(ΔP_{TH}) and p2p(ΔP_{AB}) are used in the calculation of M using equation 65.

---

4 Given that p2p(ΔP_{TH}) and p2p(ΔP_{AB}) have Gaussian distributions and also that these variables are supposed to be statistically independent, their weighted sum also adopts a Gaussian distribution. The most probable value of such a distribution is its average value.

7.6.5.3. Discussion of the QDC method

The QDC method proposed by Sackner is strongly questionable in many respects.

The first is that, in practice, it is not possible for naturally breathing subjects to breathe with an exactly constant tidal volume. It follows that the initial assumption that the zero variance population of tidal volume can never be completely true.

The second questionable aspect is that the method considers the population of $p2p(\Delta P_{TH})$ and $p2p(\Delta P_{AB})$ as statistically independent variables. It was shown by Sackner that $p2p(\Delta P_{TH})$ and $p2p(\Delta P_{AB})$ fit normal distribution curves during quiet breathing but this does not allow one to draw a conclusion concerning their statistical dependence. More importantly, if these variables were statistically independent, then they could not both have a zero variance (a fortiori with Gaussian distributions).

The third questionable aspect is that the resulting calibration coefficient M is necessarily negative, implying the plethysmographic signals are not used in a constructive but systematically destructive manner to form the estimation of ΔV. Despite the possible exactness of the result, there is no physiological ground to prove the validity of this consequence of the QDC method.

7.7. Plethysmographic respiratory monitoring instruments

The idea that respiratory measurements can be estimated from the direct observation of the human torso is not a new concept. For example, in 1905 Haldane designed a box that surrounded all but the head of the subject, The box was air tight and sealed around the neck. Changes in the volume from the box were then mapped onto changes in air respired by the subject [Haldane (1905)]. This idea is still use nowadays. Since that time, other less cumbersome non-invasive techniques have been designed. In order to measure changes in thoracic and abdominal circumferences during ventilation, Wade used rubber tubes filled with mercury and stretched around the subject's torso. These tubes were connected in series to Wheatstone bridges. Changes in torso circumference during respiration, stretched the tubes and altered the resistance offered by the mercury. These resistance changes were in turn detected as imbalances in the Wheatstone bridges. [Wade (1954)]

Konno and Mead describe the measurement of chest wall antero-posterior diameters with an instrument consisting of longitudinal transducers attached via a system of string, weight and pulley, to the subject's chest wall [Konno and Mead (1966)]. Mead also described a non-invasive monitoring device based on the application of magnetometers coils attached to the skin at opposite surfaces of the chest wall in order to estimate antero posterior diameters. One coil generated a

magnetic field, inducing a voltage in the other, the strength of which was dependent on the distance between them [Mead et al (1967)].

7.7.1. Respiratory Inductive Plethysmograph (RIP)

Respiratory Inductive Plethysmography primarily consists of estimating respiratory volume by measurement of the cross sectional area of the thoracic and abdominal compartments. Evolving from the devices presented in section 7.7, the RIP was presented as a non-invasive transducer that estimated changes in cross sectional areas, rather than circumferences or antero-posterior diameters of the thorax and abdomen [Cohn et al (1978), Cohn et al (1982), Cohn et al (1994)].

The cross sectional area measurements are obtained by encircling the thorax and abdomen with insulated electric wires. Each wire is stitched into an elasticated belt in a specific pattern so that the enveloping belt can stretch as the subject breathes. The cross sectional area variation is transduced as an inductance variation that is exploited to produce frequency variation of the output signal of an oscillator. A second and final signal, whose amplitude is directly dependent on the cross sectional area encircled by the sensor, is generated by demodulating the first signal.

Theoretical considerations of inductance plethysmography suggest that there is a slightly curvilinear relationship between the inductance of the wire and the cross sectional area enclosed by the wire. [Crawford et al (1983), Martinot-Lagarde et al (1988)]. However experimental results indicate that in the physiological range this could be regarded as a strictly linear relationship: with a near circular cross section, the error is less than 5%. [Watson et al (1988)] Martinot-Lagarde also demonstrated that the pseudo linear relationship between inductance and cross sectional area is shape dependent, i.e. it is only true for shapes that remain the same as their area changed [Martinot-Lagarde et al (1988)]. This is not so in the human ribcage and abdomen, particularly at the extremes of lung volume. Furthermore, the shape and the relative position of the inductance wire pattern of the sensor also influences the relation between cross sectional area and inductance.

The RIP can present health and safety risks as the patient is in permanent electric contact with the system during the investigation. Efficient electrical isolation from the power mains is necessary to ensure the patient's safety.

These problems are important considerations in the use of the RIP since they suggest inherent inaccuracies in the device. However, because of the absence of rival commercialised technology, the RIP has a leading position in the market for unobtrusive respiratory monitoring instruments.

At present, there are three manufactures of RIP devices, namely NIMS, Sensormedics and Vivometrics. All three companies offer RIP devices with interface caspability to a computer for data acquisition over long periods of time.

Due to its simplicity and light-weight, the RIP can allow long-term respiratory monitoring inside and eventually outside the laboratory. However, during long investigations such as sleep disorder studies, the commercialised RIP devices all have inherent problems associated with the slippage of

transducer belts, the disconnection of plugs and the breaking of cables. Additionally, patients report experiencing discomfort from wearing the two sensor belts and the multiple wires taped on their skin during such long investigation. It was concluded that the present RIP system, as commercialised by NIMS and SensorMedics, had limited ergonomics, in particular for usage in clinical settings by practitioners [Wilhem et al (2001)].

Recently, the RIP has been adapted for ambulatory respiratory measurement along with other vital sign monitors, such as an electrocardiograph, into a single garment equipped with a telemetric system for continuous remote sensing. The product has been called "Life Shirt". Its prototype was developed by the corporation Vivometrics in the U.S.A. The Life Shirt can also be purchased Europe. It has been assigned 11 patents covering the garment, the embedded sensors, and the software application dedicated to signal analysis.

The application of RIP technology in ambulatory patients is more problematic than recording supervised, resting patients in the laboratory. The plethysmographic sensors react not only to respiratory activity but also to general postural changes, bending of the trunk, or vibration of the abdomen, chest, or breasts due to accelerative forces on the body. Walking produces rhythmical fluctuations on top of the breathing signal [Wilhelm & Roth (1996)]. Another limitation of this ambulatory RIP technology at the current stage is that, although most of the extracted parameters have been validated within the laboratory, the usefulness of several of the parameters in ambulatory settings has not been demonstrated. Certain parameters may only be reliably obtainable during standardised sitting or supine resting periods [Wilhelm et al. (2001)].

7.7.2. Fibre Optic Respiratory Plethysmograph

Like the RIP, the FORP is based on the concept of ventilation volume estimation by thoracic and abdominal compartment volume estimation. In the FORP, the modality of volume estimation is by torso circumference measurement, instead of cross sectional measurement, and the circumference sensor is based on the macrobending loss effect in multimode optical fibre (see section 4.3.4). The FORP sits in the position of being an alternative approach to plethysmographic respiratory monitoring. By using an optical technique rather than electrical to estimate respiratory volume, the FORP has enhanced functionality compared to the RIP: The FORP can be used in harsh electromagnetic environments (typically the inside of a Magnetic Resonance scanner) and in underwater environments.

7.7.2.1. Historical development

The idea of a fibre optic based non-invasive respiratory monitoring was first proposed by Augousti in 1992. The development of the initial concept and its enhancement has been largely supported by Raza and these efforts been documented in detail. [Augousti and Raza (1993), Raza and Augousti (1994), Raza and Augousti (1995), Augousti (1997)] Raza is responsible for the creation of the first prototype and the transfer of the RIP calibration methods to the FORP. More importantly perhaps, Raza adapted the FORP to a single thoracic channel device capable of gating an MR scanner when image generation required an acquisition time longer than the subject's longest breath holding period. The subject inside the MR scanner is provided with visual feedback on his torso position. The information provided by the feedback helps one to adopt a stable and consistent chest position that minimises motion blurring due to chest displacement. Consequently, the scanner resolution is fully exploited and the scanner's imaging algorithm produces more detailed pictures

The development of the FORP has also seen external contributors. Davis [Davis and Mazzolini (1997)] derived a slightly different version of the FORP using light injection from a laser diode into a monomode fibre. The sensor itself contained only a single loop of fibre. The FORP proved successful at monitoring the respiratory activity of anaesthetised piglets placed in a High Frequency Oscillatory Ventilator (HFOV), otherwise known as an artificial respirator.

Using a sensor made of a single loop of multimode polymer optical fibre and an infrared LED as light source, Babchenko produced a version of the FORP that was capable of detecting chest circumference variation arising from the cardiovascular pulse [Babchenko (1999)].

7.7.3. Conceptual difference between the RIP and the FORP

Unlike the RIP, the FORP does not directly measure the torso cross sectional area but estimates it by measuring the torso circumference. Estimating a closed surface from its perimeter requires deriving an expression of the perimeter to surface area relationship. Such a relationship depends on the geometry of the surface. By assuming a circular geometry, it can be shown that the surface area variation is proportional to its perimeter, if we consider that the radius variations are negligible compared to the radius value. Augousti uses this assumption to reformulate the Konno and Mead model as the two connected cylinders model and to establish an equivalence between tidal volume estimation obtained from the FORP and the RIP [Augousti (1997)].

Such an assumption provides a straightforward perimeter to surface relation but actually corresponds to a loss of generality compared to the original Konno and Mead model. Moreover, the functional anatomy of the torso cross sectional geometry during respiration indicates that a circular area is only a particular case of a more general cross sectional geometry. A more realistic approach is to consider an elliptic geometry. This is illustrated in the figure below where an ellipse (long

dashes) and two circles (short dashes) have been represented on top of a picture showing a transverse section of the human trunk in the thoraco-abdominal region:

Figure 2: Transverse section of the human trunk, perpendicular to the plane in which the subject is lying, passing through the 12th thoracic vertebra [Koritke & Sick (1988)]. The section shows the thoraco-abdominal region. The subject is a 40 year old male. The inner circle has a diameter equal to the anteroposterior distance and the outside circle has diameter equal to the lateral distance. The two perpendicular straight lines are the long and short axis of the ellipse and their intersection represents the centre of the ellipse and the circles.

In the figure above, the ratio of long and short radii of the ellipse is 1.35. This only constitutese a particular case and, more generally, this ratio varies with the subject's position, morphology, physiological condition and muscular activity (including breathing).

A circle being a particular form of an ellipse, this new approach can be regarded as an extension of the previous one. An expression for the small surface area variations of an ellipse with respect to both radii is given by:

$$\Delta S = \pi r_1 \Delta r_2 + \pi r_2 \Delta r_1 \qquad \text{(Eq. 66)}$$

where r_1 and r_2 are the ellipse's radii, and $\Delta r1$ and $\Delta r2$ are the radii variations. In the following, we may assume that radii r_1 and r_2 are constants because the radii variations are negligible compared to r_1 and r_2. Therefore, if we divide the equation above by πr_1, we obtain an expression that is proportional to the surface area variation:

$$\Delta S \propto \frac{r_2}{r_1} \Delta r_1 + \Delta r_2 \qquad \text{(Eq. 67)}$$

The above equation expresses the relative contribution of each radius variation to the estimation of the surface area variation. This relative contribution takes the form of the ratio of r_2 and r_1. This expression is of particular importance because, if we want to estimate a surface area variation from

a perimeter variation, the perimeter variation equation must have the same layout as the equation above and the factor expressing the relative contribution of Δr_1 and Δr_2 must have a value exactly equal or similar to r_2/r_1.

One feature of the ellipse is that it has no exact expression for its circumference as a function of both radii. Different approximations exist however, and the Ramanujan equation is considered to be the most accurate, even for large differences of r_1 and r_2. Additionally, it is easy to derive, which is important to us because we are looking for the perimeter variation. The Ramanujan equation is given as:

$$C = \pi\left(3(r_1 + r_2) - \sqrt{(3r_1 + r_2)(r_1 + 3r_2)}\right) \quad \text{(Eq. 68)}$$

where C is the ellipse circumference, and r_1 and r_2 are the ellipse radii. It can be shown that an expression for small variations of the ellipse circumference as a function of both radii is given by:

$$\Delta C = \pi\left[A \cdot \Delta r_1 + B \cdot \Delta r_2\right] \quad \text{(Eq. 69)}$$

where A and B are factors given by:

$$A = 3 - \frac{3r_1 + 5r_2}{\sqrt{(3r_1 + r_2)(r_1 + 3r_2)}} \quad \text{(Eq. 70)}$$

$$B = 3 - \frac{5r_1 + 3r_2}{\sqrt{(3r_1 + r_2)(r_1 + 3r_2)}} \quad \text{(Eq. 71)}$$

Considering that expressions for A and B are functions of r_1 and r_2 and given that r_1 and r_2 have been approximated as constants, we can say that A and B can be approximated as constants. In this case, by dividing the equation 69 by πB, we can write another expression that is proportional to the circumference variation. This is written as:

$$\Delta C \propto \frac{A}{B}\Delta r_1 + \Delta r_2 \quad \text{(Eq. 72)}$$

We can clearly see that, if the ratio of A/B forms an approximation of r_2/r_1, then we are in possession of the proof that ΔC is a biased but efficient estimation of ΔS for small radii variations. To study this hypothesis, we will perform a numerical analysis of the relative difference of A/B and r_2/r_1 We define the relative error E as a function of r_1 and r_2:

$$E(r_1, r_2) = \frac{\frac{A}{B} - \frac{r_2}{r_1}}{\frac{r_2}{r_1}} = \frac{A \cdot r_1}{B \cdot r_2} - 1 \qquad \text{(Eq. 73)}$$

E is calculated for values of radii r_1 and r_2 ranging from 0.1 to 1.0 m by steps of 0.05 m. Consequently, a region of the plane formed by r_1 and r_2 where E does not exceed +/- 15% (say) can be deduced and this region is represented below:

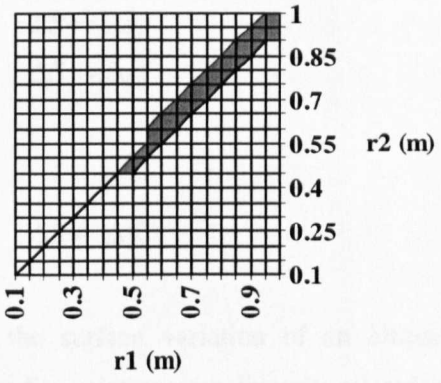

Figure 3: Ellipse radii domain for which the factor A/B determining the relative contribution of Δr_1 to Δr_2 is accurate to within 15%. Each cell has dimension 5cm by 5cm. The greyed region shows where absolute value of E is smaller than or equal to 15 %.

Graphical interpretation of the figure above shows that the limit of the validity of the factor A/B occupies a narrow band increasing with r_1 and r_2 and positioned around the straight line defined by $r_1 = r_2$. This indicates that the correct value of A/B can only be obtained for an almost circular cross sectional geometry. In particular, E is zero when $r_1 = r_2$, i.e. when the ellipse is a circle. In this case, the surface area variation would be perfectly estimated by the circumference variation. For all other cases, the circumference measurement is a distorted surface area estimation. By distorted estimation, we understand this to mean that the shape (regardless of the scale and position) of ΔC does not match the shape of ΔS because the factor A/B has an erroneous value. As we do not have access to the value of Δr_1 and Δr_2 during the circumference measurement, the estimation error is intrinsic: ΔC cannot be manipulated to replicate ΔS.

This is not as bad as it seems if we consider that Δr_1 and Δr_2 are not statistically independent variables[5].

Considering that the ratio of r_1 and r_2 is constant, it seems a fair assumption to suppose that the relation between Δr_1 and Δr_2 is of a linear nature. Having Δr_1 and Δr_2 linearly related effectively corresponds to a reduction of the number of degrees of freedom from 2 to 1 in the problem of

5 This can be demonstrated simply by breathing while placing our hands on one's the torso; Δr_1 and Δr_2 increase and decrease together.

estimating ΔS from Δr_1 and Δr_2. The expression of small circumference variation is now simplified. It is given by:

$$\text{Or} \quad \begin{matrix} \Delta C \propto \Delta r_1 \\ \Delta C \propto \Delta r_2 \end{matrix} \qquad \text{(Eq. 74)}$$

In the same manner, we can now write:

$$\text{Or} \quad \begin{matrix} \Delta S \propto \Delta r_1 \\ \Delta S \propto \Delta r_2 \end{matrix} \qquad \text{(Eq. 75)}$$

Finally, it follows that:

$$\Delta S \propto \Delta C \qquad \text{(Eq. 76)}$$

This expression signifies that the surface variation of an ellipse can be estimated from its circumference variation if the radii variations are linearly related and if the radii variations are negligible compared to the initial radii values. This ellipsoid model is applied to the torso cross section to demonstrate that the FORP can produce measurements equivalent to the RIP, during quiet breathing.

On table 1 below is represented the thoracic and abdominal circumference variations for 5 different subjects during exaggerated breathing. During normal breathing, these variations rarely exceed 1 cm.

	Females		Males	
Subjects	Thorax (cm)	Abdomen (cm)	Thorax (cm)	Abdomen (cm)
1	4.5	6	7	8
2	7.5	7	6	6
3	5	6	6	7
4	8.5	8	4	6.5
5	4.5	11	9	8
Average	6	7.6	6.4	7.1
SD	1.87	2.07	1.81	0.89

Table 1: Maximum torso excursion during exaggerated ventilation. The largest excursion is about 7 to 8 cm. It is important to note that the average torso excursion during quiet breathing is about 1 cm. [Raza (1998)]

7.7.4. Design and achievements of the FORP

The chest circumference measurement in the FORP is carried out by an optical fibre sensor. The use of optical fibres for plethysmography, with particular application to respiratory monitoring has demonstrated the feasibility of using the macrobending loss effect (MBLE) produced in fibre coils for sensitive measurements of chest circumference. The sensor principle is to arrange the optical fibre in a configuration that converts a variation of elongation applied to the sensor into a variation of fibre optic loop radius, which in turn modulates the optical transmission characteristics of the fibre by means of the MBLE: As the fibre loop experiences increasing curvature, the optical power transmitted by the fibre is altered by the eradication of high order modes. This effectively results in a lowering of the total mode power emerging from the fibre end.

Two separate approaches have been proposed so far for the structure of the sensor: 1) multiple coil structure or 2) single coil structure. These sensors, as well as their optronic and acquisition systems are now reviewed below.

7.7.4.1. FORP prototype by Raza and Augousti

7.7.4.1.1. Sensor

The FORP prototype by Raza and Augousti was based on a multiple coil sensor [Augousti and Raza (1993)]. This sensor consisted of a multiple optical fibre coil configuration mounted on an elasticated bandage, as can be seen below on figure 4.

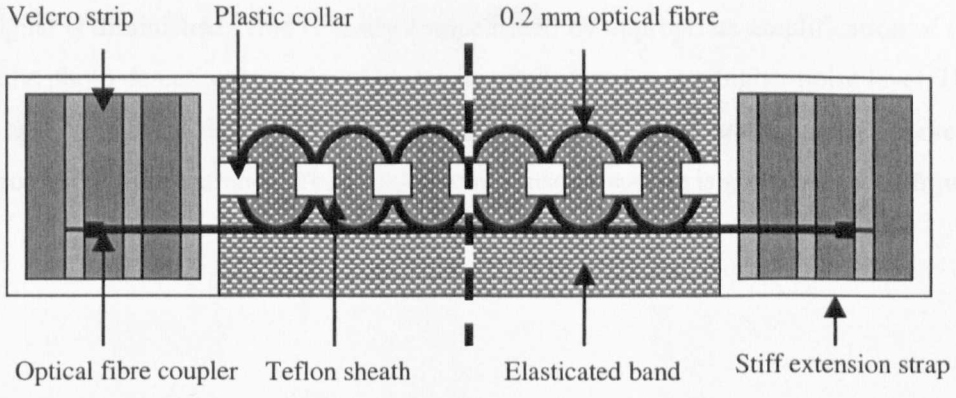

Figure 4: The multiple coil sensor used by Raza

The elasticated bandage was created by sewing together a number of sections of Tubigrip support bandage to form a single strip of material, capable of stretching in the longitudinal dimension of the bandage length. To ensure that the sensor band hugged the torso, two sections of band were sewn on top of each other. Rather than sewing the strip into a permanently closed loop, two strips of rigid material (usually used to support collars in shirts), were sewn into either side of the band. Strips of Velcro were then attached by fabric glue either end of the band strip. This allowed the sensor band to be worn by a variety of test subjects with differing build. The fibres were held in place using

Teflon coated plastic collars, which forced them to take up a two-dimensional coil configuration. The size of the loops was carefully chosen to optimize the sensitivity of the sensor whilst maintaining an adequate signal strength: too large and the sensor would be insensitive to variations in the loop radius; too small and the sensor elongation range would be restricted. To prevent the fibre from retracting when the band expanded, two sections of tight plastic tubing acting as stops were placed on the two ends of the fibre that protruded out of the band. Because the sensor was prone to damage from overstretching, a band of collar stiffening material was sandwiched in between the two halves of the elasticated band, with the two ends of the material stitched down behind the two terminal ends of the optical fibre. This ensured that the band could not be overstretched. The optical fibre configuration did not involve any tying of the fibres and was therefore mathematically equivalent to an unknot. The tendency of the fibre would be to come back to its naturally untwisted state if the plastic collars did not restrain it, thereby leading to the production of forces transverse to the plane of the coil. These forces lead to increased friction between the coils and the collars, which in turn limits sensor resolution. This configuration is mechanically complex in terms of mounting the fibre and the collars on the band because it contains multiple coils that are in a condition of forced equilibrium.

The friction was minimised in this configuration by reducing the fibre stiffness by using thinner fibres, typically with a diameter of 0.2 mm. This raises two issues: firstly, coupling light into small diameter fibre is delicate because of the possible alignment errors between the fibre ends and the light source and detector; secondly, a smaller core diameter restricts the maximum possible optical power being carried by the fibre. Both issues contribute to the same effect, namely the output optical signal is diminished. This is easily compensated by appropriate amplification of the optical signal at the photo-detection stage but this also inevitably produces a higher noise level. Ultimately, the most appropriate solutions are firstly to use a larger optical fibre and secondly a more powerful light source at the fibre entrance. The multiple coil sensor response is given below on figure 5.

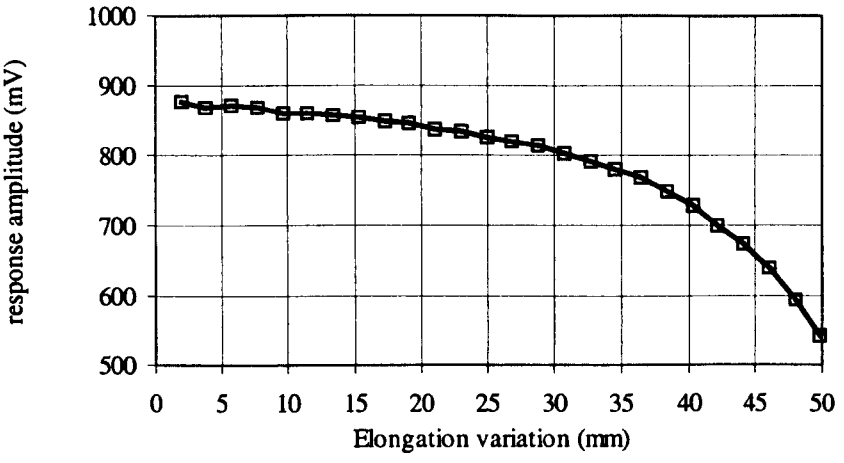

Figure 5: The multiple coil sensor response. The response curve is essentially non-linear. A linear approximation over a 5 mm range can however be considered as acceptable. The response curve contains non-monotonic portions which are attributed to the complex mechanical assembly of the fibre. Additionally, the sensor response has a restricted magnitude: in the graph above, the working amplitude range offers only 300 mV difference between no elongation and maximal elongation and only 50 mV between no elongation and half the maximal elongation.

The multiple coil structure has the important property of being adaptable to any extension range by adding or subtracting coils at will during sensor construction. This renders the FORP versatile for any measurement range.

A numerical simulation of the sensor optical transmission ratio as a function of the elongation applied to it and the number of coils was undertaken. It shows that, for a given elongation, the sensor sensitivity decreases with an increasing number of coils. The multiple coil sensor simulation graph is shown in figure 6 on the next page.

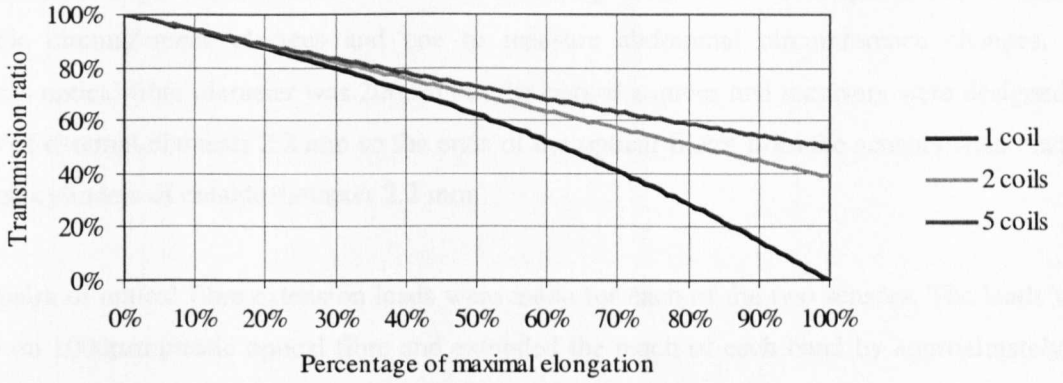

Figure 6: Multiple coil sensor simulation. The "1 coil" response curve was obtained from the quadratic regression of a real single coil sensor. The regression had a Pearson coefficient value of 0.9999, indicating a nearly perfect correlation and providing confidence in the simulation. The regression equation was extended to 2 and 5 coils by assuming equal distribution of the sensor elongation between all coils and by raising the transmission ratio to the power of the number of coils. In other words, if $T(\Delta x)$ represents the transmission ratio of the single coil sensor, then $T^N(\Delta x/N)$ represents the transmission ratio of the N coils sensor. It can be seen that the accumulation of coils in series does not contribute to increasing the sensor linearity.

Theoretically, if all coils have perfect and equal elasticity[6], then the elongation applied to the sensor is equally shared among the coils. For example, in a three coil configuration, each coil experiences one third of the global elongation applied to the sensor.

In practice this is not the case: as mentioned above, the mechanical complexity of the sensor generates frictional forces which hinders the extension of the optical fibre coils. It follows that, during extension, different coils experience different elongations. Additionally, as the extension is released the coils may retract in a different sequence other than the reverse of that which occurred when they expanded. The consequence of these two effects is to reduce the monotonicity of the overall sensor response and to increase its hysteresis.

From these observations, we concluded that a multiple coil structure was not necessarily well adapted to the precise measurement of elongation. Given the relatively restricted range of elongation to be measured in respiratory plethysmography[7], it was found that a single coil structure seemed sufficient to cover most torso circumference variations.

---

6 An elongating structure is said to be elastic if its elongation is proportional to the force producing it.

7 The average of the maximum torso excursion during exaggerated ventilation is about 7 to 8 cm. The average torso excursion during quiet breathing is about 1 cm [Raza (1998)]. See table 1 in section 7.7.3.

7.7.4.1.2. Sensor operation and acquisition system

Two identical copies of the sensor were built according to the above description: one to measure thoracic circumference changes and one to measure abdominal circumference changes. The sensor's optical fibre diameter was 200μm but the optical sources and receivers were designed for fibers of external diameter 2.2 mm so the ends of the optical fibers from the sensors were encased in brass cylinders of outside diameter 2.2 mm.

Two pairs of optical fibre extension leads were made for each of the two sensors. The leads were based on 1000μm plastic optical fibre and extended the reach of each band by approximately 1.5 meters. A suitable connection system was required to ensure proper continuity in optical signal conduction between the sensor fibres and the extension lead fibres. Additionally, it was required that this connection could be undone to adapt different extension leads or different sensors if required.

Most techniques to connect/disconnect optical fibres rely on expensive precision connectors. An inexpensive connection system was achieved by altering the plastic casing of the SIEMENS SFH series of LEDs and photodiodes. The optronic elements were drilled out and the back end of the cases were chopped off. A brass tube of inner diameter 2.2 mm was employed to align two drilled out plastic cases back-to-back and the two cases were glued together. This simple connector permits the joining of optical fibres of outer diameter 2.2 mm such as the 1 mm plastic optical fibres. The benefit of employing a plastic casing is to take advantage of the built-in screw mechanism.

To ensure good continuity of the optical signal at the optical fibre junction, a coupling gel matching the refractive index of the fibre core was used. Below is shown a multiple optical fibre connector: four pairs of back to back plastic casings have been grouped into an aluminum holder to permit the connection of two FORP sensors to two pairs of extension leads.

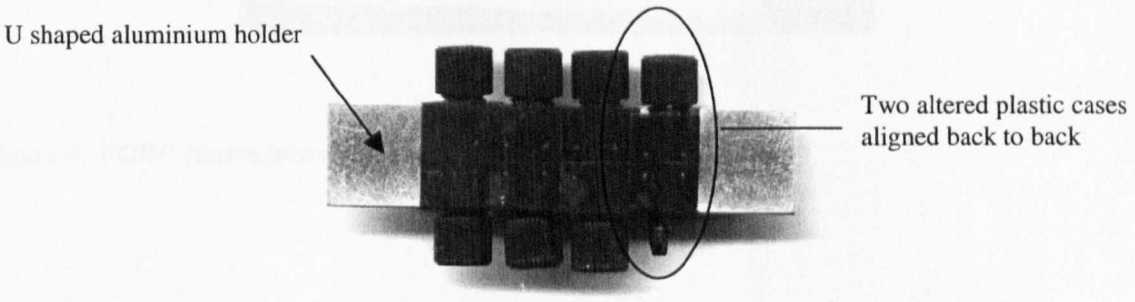

Figure 7: Multiple optical fibres connectors

At the optical end of each sensor was an infrared light source (SIEMENS SFH450, λ= 950 nm). The wavelength matched the fibre transmission spectrum and on the other end of the fibre was a photodiode corresponding to the light source (SIEMENS SFH250V). The sources and the detectors are housed in a plastic casing that provides a "screw in" attaching port for optical fibre of 2.2mm outside diameter. The sources are continuously powered. The receivers are wired to two independent current to voltage converters and amplifiers. Both signals were fed into a PC based 10 bits acquisition system for processing and storage. Due to the small peak to peak amplitude of the plethysmographic signals, a large amount of amplification was required to fit the signals to the input range of the analogue to digital converter and consequently the signals contained a large proportion of noise. The acquisition system itself consisted of a multiple analogue inputs external module that was connected to a PC via the parallel port. Three inputs were used: one for the signal originating from the thoracic sensor, one for the signal originating from the abdominal sensor and one for the spirometer signal.

The advantage of exploiting the parallel port is that the acquisition system could be used with absolutely any PC, either desktop or laptop, therefore providing excellent portability of the system. In practice, a laptop was employed during investigations at different locations. To generate binary data files during acquisition and interpret them, either offline or in real time, data acquisition and analysis software were written in Visual Basic for MS-DOS and later for Windows.

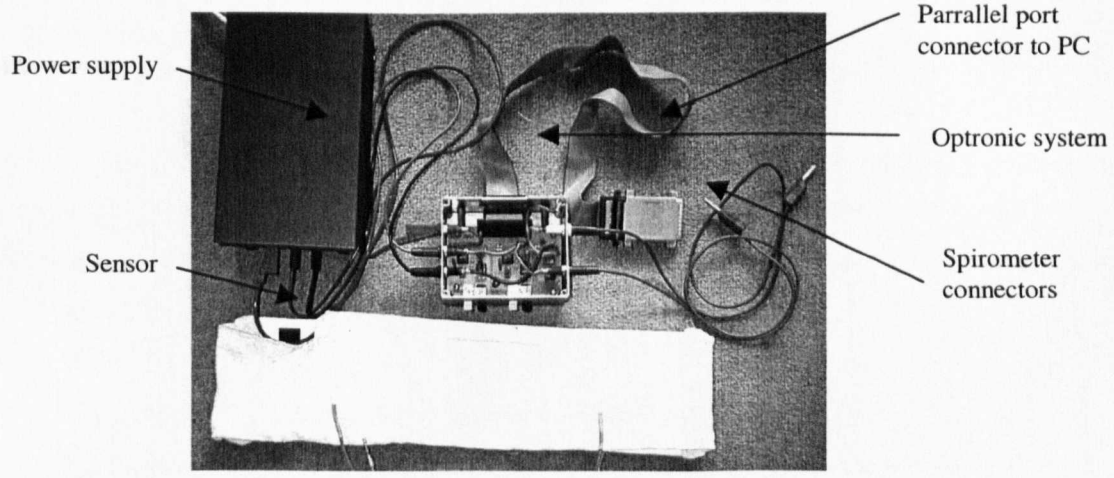

Figure 8: FORP system prototype by Raza and Augousti

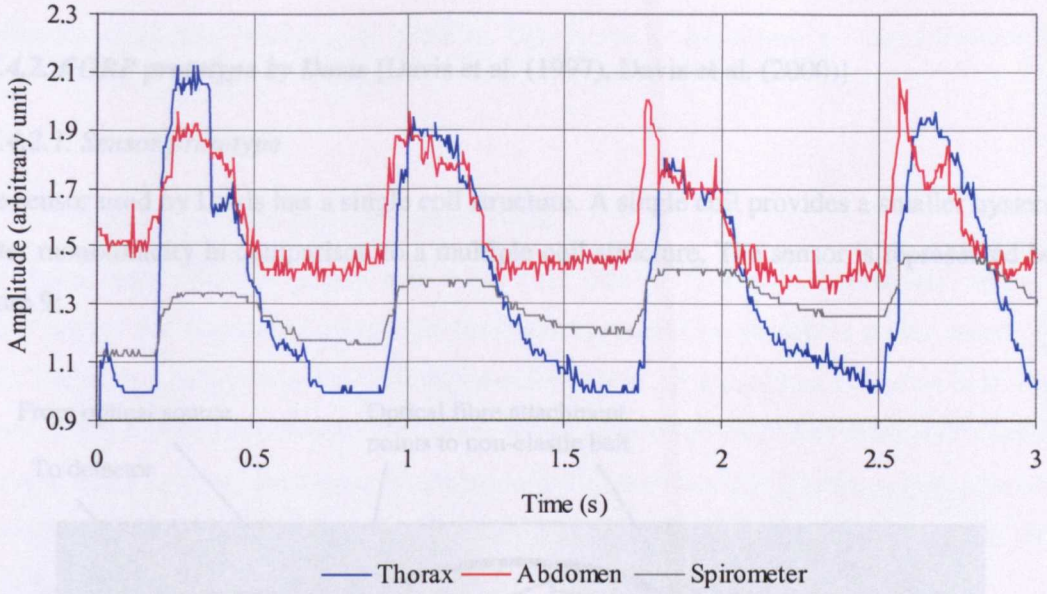

Figure 9: Typical FORP signals acquired with Raza's prototype while the subject was breathing normally. None of the signals have been filtered. The prototype's low amplitude resolution and low temporal resolution is clearly seen here. Note in particular that the spirometric signal is ergodic, whereas the plethysmographic signals are not really ergodic: This is difference is purely due to instrumental errors. More importantly, note the difference of general aspect between these signals and the FORP signals presented in section 11.2.2 on figures 2 to 5: The later FORP signals are ergodic.

7.7.4.2. FORP prototype by Davis [Davis et al. (1997), Davis et al. (2000)]

7.7.4.2.1. Sensor prototype

The sensor used by Davis has a single coil structure. A single coil provides a smaller hysteresis and better monotonicity in comparison to a multiple coil structure. The sensor is represented below on figure 9:

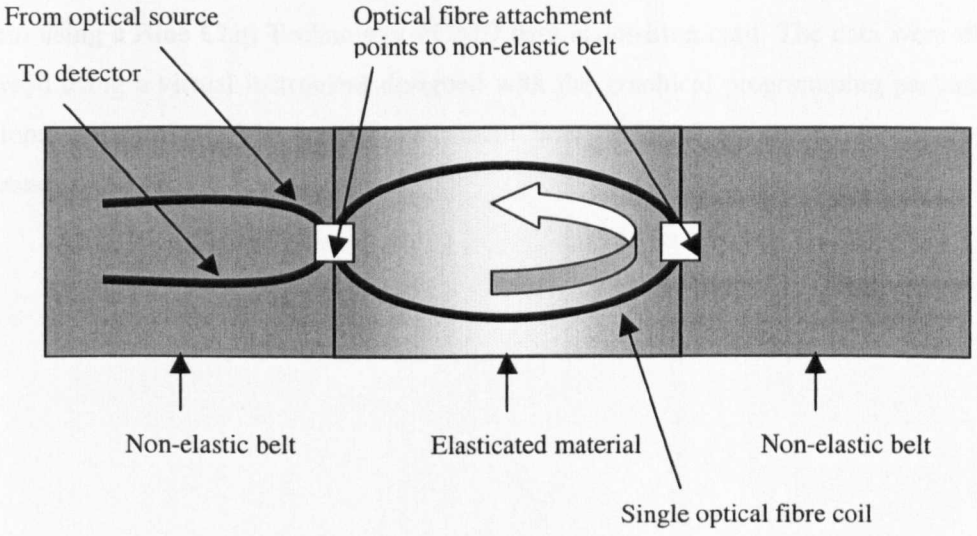

Figure 10: The single coil sensor used by Davis. The three dimensional arrow represents the direction of propagation of light in the optical fibre.

Two identical sensor belts were created. Each belt was constructed to be inextensible apart from an elastic expansion strip (approximately 3 cm in length) into which a single fibre loop was stitched. This ensured that the total change in thorax and abdomen circumferences occurred over a limited region which maximised the curvature change of the loop.

A feature of this sensor is to use a monomode optical fibre of diameter 5.6 μm. The choice of a monomode fibre helped in producing better measurement repeatability, compared to a multimode fibre. This could be due to the high sensitivity of the MBLE on mode population: multimode fibres might increase the variability of the mode population.

The sensor response was safely estimated by linear regression for an extension domain range of only 6 mm, the Pearson correlation coefficient in this case was equal to 0.9932. It is presumed that the use of a monomode fibre rendered the sensor fragile, which severely limited the elongation range.

The use of such a sensor is not well adapted to respiratory plethysmography of adults and Davis restricted her investigations to the respiratory monitoring of piglets in the context of high frequency

oscillatory ventilation [Davis et al. (2000)]. This is due to the sensor's limited linear elongation range.

7.7.4.2.2. Sensor operation and acquisition system

The sensor input was illuminated with a 780 nm, 2.86 mW GaAlAs laser diode source and the sensor output was monitored using a photodetector mounted on an optical power meter (Newport 835). The analog output from the optical power meter was interfaced to a PC based data acquisition system using a Blue Chip Technology PC30D data acquisition card. The data were displayed and analysed using a virtual instrument designed with the graphical programming package LabVIEW (National Instrument). The virtual instrument was designed to sample the analog data at a frequency of 10 Hz.

7.7.4.3. FORP prototype by Babchenko [Babchenko et al. (1999a), Babchenko et al. (1999b)]

7.7.4.3.1. Sensor prototype

The single coil sensor used by Babchenko et al. revealed a design somehow different from the one used by Davis, as represented below:

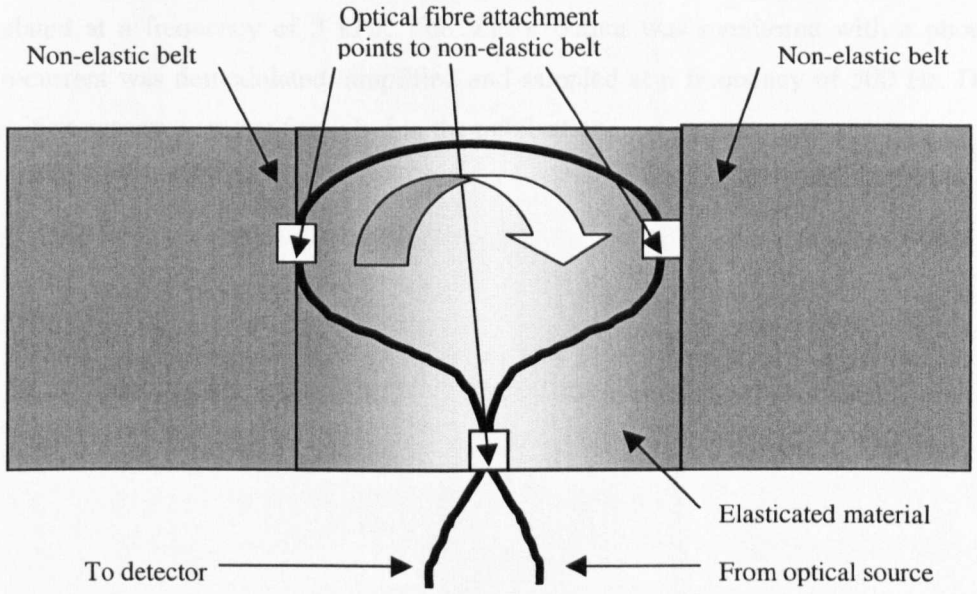

Figure 11: The single coil sensor used by Babchenko. The three dimensional arrow represents the direction of propagation of light in the optical fibre.

The sensor used by Babchenko is based on a polymer optical fibre of diameter 0.4 mm. The chest belt was divided into two sections: a short elastic part which can be lengthened during respiration, and flexible longer part of constant length. The fibre, bent approximately to a circle of about 40 mm in diameter was connected at two points to the edge of the flexible non-elastic section so that the diameter increased during inhalation.

It is reported that the sensor response curve has a curvilinear behaviour over a maximum elongation range of 5 cm. The deviation from the sensor output from linearity is smaller than 10% in the range 2 to 3 mm. The sensor response curve shows an hysteresis loop of maximal amplitude equivalent to a 0.25 mm elongation. It is assumed that the large amplitude hysteresis loop is partially due to the use of plastic optical fibres. Such fibres do indeed have a restricted elastic domain and a large plastic[8] domain.

8 The plastic domain corresponds to the elongation range where the fibre does not necessarily recover its original shape and dimensions after elongation, unlike the elastic domain. Polymer fibres usually have the mechanical property of slow relaxation, which considerably restricts their elastic domain, and therefore contributes towards hysteresis.

It was concluded that the single coil sensor used by Babchenko provided a sufficient elongation range but that its utility was restricted by 1) the limited linear elongation range and 2) the large hysteresis loop.

7.7.4.3.2. Sensor operation and acquisition system

The sensor input was illuminated from an IR LED source of wavelength 0.845 μm, which was modulated at a frequency of 3 kHz. The sensor output was monitored with a photodiode. The photo-current was demodulated, amplified and sampled at a frequency of 500 Hz. Details of the acquisition system were not furnished in the publication.

Chapter 8 PLETHYSMOGRAPHIC CARDIAC MONITORING

8.1. Introduction

Plethysmography is general technique to perform an indirect measurement or an estimation of a body volume or, by extension, a body volume variation. The cause of volume change can be diverse: In the case of the FORP for example, we have primary employed plethysmography to estimate respiratory volume from thoracic and abdominal circumferences variations due to the ventilation system. However, ventilation is not the only function responsible for torso volume changes.

The beating heart produces local volume variation in the thorax. Also, the heart emits a cardiovascular pressure pulse that propagates along the network of arteries and veins throughout the body. Such pulse also generates local body volume changes that can eventually be transmitted to the surface of the skin. For example, it is well known that cardiac pulsation can easily be picked up on the wrists or the neck simply by sensing the skin deformation with a fingertip.

Plethysmographic cardiac monitoring is the recording of local body volume variations due the activity of the cardiovascular system with a plethysmographic instrument.

Because a plethysmographic method cannot differentiate the phenomena that lead to body volume variations, both respiratory and cardiovascular activity can be perceived with the same plethysmographic instrument, with the difference that cardiovascular activity usually results in plethysmographic signals of much smaller amplitudes than respiratory ones.

In this chapter, we will review plethysmographic methods that permit interrogation of respiration and cardiovascular activity.

8.2. Impedance plethysmography

Impedance plethysmography is an estimation of torso volume variation by measurement of electrical impedance variations of the torso. The impedance change is generated by the ventilation system (air flow) and/or by the cardiovascular system (blood flow). Therefore, as all body plethysmographic techniques, it permits the simultaneous interrogation of respiratory and cardiac activity. Respiratory monitoring with impedance plethysmography is called impedance pneumography (IPG) [Allison (1964), Hamilton (1967)]. This technique is detailed in section 2.2.2.4. Cardiac monitoring with impedance plethysmography is called impedance cardiography (ICG) [Patterson et al. (1964), Kubicek et al. (1966), Nyboer (1970)]

ICG is primary a technique to estimate cardiac output. A sinusoidal current of amplitude 3 to 6 mA and frequency about 100 kHz is applied on the skin across the torso (from the neck to the thoraco-abdominal region). The resultant voltage is measured approximately at the points of current injection. An impedance signal is generated by analysing the phase and the amplitude of the voltage over current ratio. Based on the observation that blood is capable of current conduction, it was reasoned that blood volume variation in the region of ICG interrogation would participate in

producing impedance variation. Indeed, the impedance signal varies at the frequency of the heart rate. Deriving a quantitative estimation of stroke volume from the impedance signal requires to address the problem of how and how much electrical currents distribute across the torso. This problem is complex and incompletely understood so the ICG technique relies principally on simplified models of the torso's electrical conduction. Despite giving satisfactory results in the laboratory, interpretation of measured data remains controversial in clinical conditions because simplification of the torso conduction model can be perceived as unrealistic and has hampered the acceptance of ICG as a clinical method to estimate cardiac output. [Kauppinen et al. (2001)]

8.3. Electrical Impedance Tomography

Electrical Impedance Tomography (EIT) is an extension of the concept of Impedance Plethysmography where not two but up to sixteen electrodes are being used to circulate low currents in the body. An equal number of electrodes are used to measure the impedance variations around the subject and clever algorithms are developed to construct a 2D or 3D image from these measurements of impedance. EIT was already mentioned in section 2.5.2.5.

8.4. Photoplethysmography

In photoplethysmography (PPG), blood volume variation is estimated from the measurement of light reflected by the inner layers of the subject's skin. The intensity variations of the measured light arise from vascular changes in the tissue bed that are synchronous with the arterial pressure pulsation [Challoner (1979), Almond and Cooke (1989)]. However, other slow variations in blood flow also affect the modulation of light intensity (Kamal et al. 1989, Dorlas and Nijboer (1985)]. One of these variations is the venous return to the heart, due to change in intrathoracic pressure caused by respiration. In other words, this technique can be employed for simultaneous monitoring of cardiac and respiratory activity.

The sensor head is made of two optical fibres, one to guide the light that illuminate the skin and another to receive the light being reflected by the skin. Because the sensor head can be made of non-metallic components, this technique has the advantage to be transparent to EM field variations. For this reason, the technique has been employed in the MR scanner as an alternative to ECG gating during cardiovascular imaging. This technique was not as reliable as the ECG gating technique, presumably due to the location of the sensor on the subject's fingertip, making the signal morphology dependent.

8.5. Plethysmocardiography

The term plethysmocardiography (proposed by the author) designates the estimation of body volume changes due to cardiovascular activity by measurement of body dimensions variations. In other words, the plethysmocardiogram (PCG) is a signal depicting cardiac activity acquired with a plethysmographic instrument such as the RIP [Sackner et al. (1991)] or the FORP [Babchenko et al. (1999a), Maletras et al. (2001b)].

In particular, the thoracocardiogram (TCG) is a PCG signal obtained by a single plethysmographic transducer placed transversely at the level just below or at the xiphoid process (have a look on figure 2, section 3.2.1) and records chest wall movement resulting from heart activity. The particularity of this location is that it coincides with the level of the left ventricle and it follows that chest deformations measured at this location by plethysmography are mainly influenced by the left ventricle contractions. The physiological interpretation of the TCG signal has been thoroughly researched and reported and its main investigator is M. A. Sackner. Working along with Sackner, it was suggested by Bloch that, given the resemblance between the ventricular activity signal provided by the TCG and the ventricular volume curves obtained by echocardiography, the TCG could be recognised as a qualitative measure of the left ventricular volume curve. More precisely, in 170 comparisons of beat-by-beat stroke volumes, 89% felt within ± 20 % of echocardiographic estimates. [Bloch et al. (1998)]

Being able to monitor the left ventricular activity is of paramount importance, given that this ventricle is responsible for providing oxygenated blood to the whole body (except the lungs). In fact, thoracocardiography is the only simple technique that permits to monitor heart movement non-invasively.

Under resting conditions, the amplitude of the TCG only occupies 3 to 5 % of the respiratory signal amplitude so a better quality TCG signal is obtained by holding breath during plethysmographic examination. Depending on the subject's health however, breath holding is not always an applicable strategy. It results that one must dispose of an adequate band pass filtering technique to extract the TCG from the plethysmographic signal while breathing. At this stage, this is realised by an off line adaptive frequency digital band pass filter: The measure of the heart rate (derived from a simultaneously acquired ECG signal) is used to update the cut-off frequencies of the digital band pass filter.

Despite the repeatability of the TCG, the problem of extracting it from the respiratory trace is intrinsically linked to the behaviour of the ventilation activity. The latest is in essence very dispersive because there is no definite range of acceptable values of respiratory activity. There are number of reasons for this: First, respiratory activity is multi-dimensional in nature because derived parameters like depth, rate, and pauses can vary independently from each other. The respiratory waveform can therefore assume a wide variety of patterns. Second, respiratory activity can vary

profoundly from breath to breath. Third, speech is accompanied by complicated variable breathing patterns. Fourth, voluntary control of breathing can produce any variety of waveforms.
All these factors contribute to make the recognition of the TCG quite difficult because its spectral domain is overlaid by thus continuously changing of the respiratory signal.
To limit the dispersion of the respiratory trace (hence to facilitate the extraction of the TCG), the subject is asked to control his/her respiration. Assistance in this matter is provided by visual feed back from the respiratory trace.

Other PCG signals can be obtained by positioning a plethysmographic sensor around the neck of the subject [Jordan et al. (1984)], around the abdomen (in this case the PCG signal could be termed abdominocardiogram) (ACG)) [Maletras (2001)], or even around the head (craniocardiogram (CCG)).
However, these PCG signals do not see heart movements but local deformations of the cardiovascular system.
Sackner and other investigators of the PCG used a RIP and consequently the application of the technique was limited to environmental condition free of EM disturbance. We have seen in chapter 7 that the FORP and the RIP are equivalent systems in terms of what they measure and this implies that the FORP is equally capable of measuring TCG signals. The difference however is that the FORP can be used in condition of harsh electromagnetic environment and, in particular, in a MR scanner. [Raza et al. (1995)] This condition signifies that the FORP could be employed to obtain PCG signals that have the potential to replace ECG gating in the MR scanner. This is more thoroughly discussed in chapter 12.

Chapter 9 SENSOR DEVELOPMENT

9.1. Introduction

From the observations gathered in chapter 7 concerning the previous attempts to measure elongation with an optical fibre sensor based on MBLE, we concluded that a multiple coil structure was not necessarily the best arrangement for accurate and repeatable measurement of elongation. Such a structure generates in essence a response curve that cannot be linear. Additionally, it is prone to have a non-monotonic behaviour because of the uneven distribution of elongation among the coils during stretching and relaxation.

In an attempt to simplify the mechanical structure of the sensor, the respiratory plethysmographic sensors of Babchenko and Davis consist only of a single coil structure. [Babchenko et al. (1999a), Davis et al. (1997)] The advantage of this configuration is to present virtually no frictional effect of the optical fibre on itself. However, both sensors have a limited mechanical extension range (a few centimetres) and also a limited linear extension range (typically 6 mm). These sensors have been previously discussed in section 7.7.4.2 and section 7.7.4.3.

While retaining the idea of a non-multiple coil structure to preserve mechanical simplicity, sensor improvement would consist of designing a coil structure that provides a mechanical extension range larger than 10 centimetres to accommodate extreme breathing styles (typically during physical effort). Additionally, the new sensor would provide a linear response over a wider extension range. It would also be desirable to make the coil movement continuous, thereby limiting microscopic "stictional" effects (where the coil expands or contracts in a series of jumps, rather than by smooth movement) and to decrease the sensor hysteresis. These advantages are embodied through the use of a figure-of-eight design that is described in more detail below.

9.2. Figure-of-eight configuration

The improvements suggested above were put into practice with the redevelopment of the sensor. The transducer is comprised of a single knot, looking like a figure-of-eight.

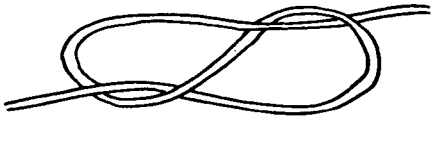

Figure 1:The figure-of-eight configuration.

As one or both ends of the fibre are pulled, the elongation variation is partially transmitted into a negative variation of curvature radius of the circular lobes on each side of the figure-of-eight. This curvature radius change implies an optical output power decrease by means of MBLE.

A silica/polymer optical fibre of core diameter 0.4mm was used in the construction of the new sensor. The fibre diameter was chosen as a compromise between the sensitivity of the attenuation with curvature and mechanical stiffness. Stiffness is required to provide the recoil force that reduces the hysteresis.

A combined visual and mechanical inspection of the figure-of-8 using the previously mentioned optical fibre permitted to understand its behaviour under a stretching action:

The figure-of-eight has a central point of symmetry and a two-dimensional geometry. Since it has only two linked coils, this configuration has numerous mechanical advantages. Lower friction at the points of contact allows the use of a larger fibre core diameter than the multiple coil sensor, which leads to a better output signal level. The sensor response is now truly monotonic and its deviation from a linear approximation over a limited elongation range is significantly reduced. See figure 9 in section 9.4.3.

This mechanical simplicity also confers robustness and reliability, as it is less prone to fibre breakage. Even more importantly, when stretched the figure-of-eight behaves rather like a spring by providing a recoil force on each of its extremities. This recoil force arises from the torque produced by the two circular arcs of the fibre. This recoil force offers the useful advantage of actively minimising the sensor hysteresis by reducing the effect of friction at the fibre crossing points. This provides excellent repeatability and reproducibility. Also, symmetric forces interacting at both fibre crossing points generate planar stability, eliminating a further source of friction. Once stretched, the transducer remains self-sustained in a plane and a guiding structure is no longer required. This makes it relatively easy to mount on a band, since the fibre needs only to be tethered at both ends, as can be seen on figure 2. Two plastic cubes are used for this purpose. On each cube, a screw allows one to preset the sensor length. Soft and robust skin compliant polymer rubber (Neoprene) is used as a mounting band. It has the same elastic properties as the Tubigrip band used in the multiple coil sensor but with the advantage of greater longevity, through the absence of any cutting and sewing operations. The fibre was connected via 2 brass cylinders to the optical source and detector, both housed in plastic, for which external and internal diameters were tailored to match the inside diameters of the plastic housings and the outside diameter of the fibre. Once fitted with these diameter converters, both fibre ends were polished to maximise the planarity of the optical junctions. The fibre was then tied into a figure-of-eight knot and mounted on the neoprene band.

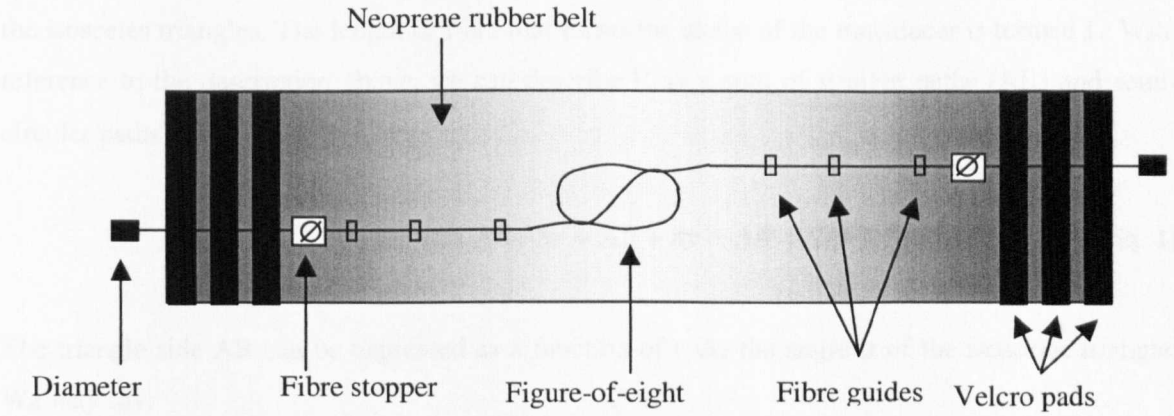

Figure 2: The single coil figure-of-eight sensor

9.3. Geometrical model of the figure-of-8 coil

A simple geometrical model of the figure-of-eight coil has been derived which has met with success in explaining the observed results. The model assists in understanding the behaviour of the radius of curvature as a function of the elongation applied to the coil. It is based on a simplification obtained by decomposition of the fibre path inside the figure-of-eight into a collection of segments that are either circular arcs or straight lines. Figure 3 represents the superimposition of the actual fibre path and the decomposition of this path, whereas figure 4 shows the model only, based on the decomposition seen in figure 3.

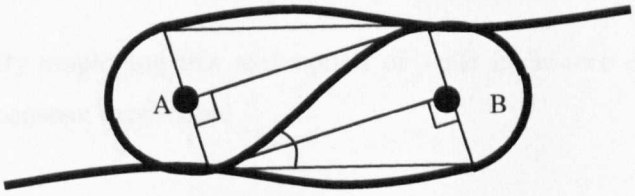

Figure 3: Geometrical decomposition of the figure-of-eight coil

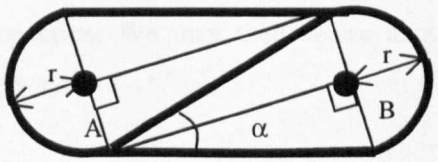

Figure 4: Model of the figure-of-eight coil

Two identical isosceles triangles and two semicircles of diameter equal to the base of the triangles are used. Note that the diameters of the two semicircles are parallel. The figure-of-eight has a central point of symmetry and therefore it can be divided into two identical substructures, shaped

like pendant drops. Each drop is constituted of a semicircle of diameter equal to the base of one of the isosceles triangles. The length of fibre that forms the shape of the transducer is termed L. With reference to the description above, we can describe L as a sum of straight paths (AB) and semi-circular paths (π r)

$$L = AB + \pi r + AB + \pi r + AB = 2\pi r + 3AB \qquad \text{(Eq. 1)}$$

The triangle side AB can be expressed as a function of r via the angle α of the isosceles triangle. We may say:

$$AB(r) = \frac{r}{\sin(\alpha/2)} \qquad \text{(Eq. 2)}$$

By visual inspection one may notice that the figure-of-eight coil appears to be scale invariant. In other words it suffers only minimal qualitative alteration in its shape while undergoing stretching, and for this reason the ratio of width to height may be assumed to be constant. According to the model, the width and height are given by:

$$W = 2r + AB \qquad \text{(Eq. 3)}$$

and

$$H = AB\sin(\alpha) \qquad \text{(Eq. 4)}$$

By employing this assumption of scale invariance of the model, we may obtain the following constant expression:

$$\frac{W}{H} = \frac{1}{\sin(\alpha)} \cdot \left(1 + \frac{2r}{AB}\right) \qquad \text{(Eq. 5)}$$

One consequence of scale invariance is that the angle α remains constant. This is confirmed by visual inspection. We may then derive another constant expressing the scale invariance in the following way:

$$K = \frac{3AB}{2\pi r} \qquad \text{(Eq. 6)}$$

The constant K here represents the ratio of the sum of linear paths to the sum of curved paths in the shape. According to equation 2, the constant K can be expressed as:

$$K = \frac{3}{2\pi \sin(\alpha/2)} \qquad \text{(Eq. 7)}$$

Using the constant K, we can now simplify the expression for L by combining equations 1 and 6:

$$L = 2\pi r(K+1) \qquad \text{(Eq. 8)}$$

Equation 8 represents a direct relationship between the radius of curvature of each semicircle and the overall length of the optical fibre transducer. The aim of this model is to express the radius r in terms of the quantity Δx, the elongation applied to the sensor since Δx, unlike L, is directly measurable. One may therefore define the total length of fibre in the figure-of-eight sensor as L_{max}, where L_{max} is defined as the sum of L and Δx, with L the length of fibre in the transducer itself and Δx being the elongation applied across both ends of the fibre.

$$L_{max} = L + \Delta x \qquad \text{(Eq. 9)}$$

Using equation 8, L_{max} is defined in the same manner as L but for a fixed radius r_{max} and we may thus write:

$$L_{max} = 2\pi r_{max}(K+1) \qquad \text{(Eq. 10)}$$

An expression for Δx may be obtained by combining equations 8, 9 and 10:

$$\Delta x = L_{max} - L = 2\pi(r_{max} - r)(K+1) \qquad \text{(Eq. 11)}$$

The quantity $r - r_{max}$ is termed Δr and we now have an expression for the change in the radius of curvature as a function of elongation:

$$\Delta x = -2\pi\Delta r(K+1) \qquad \text{(Eq. 12)}$$

The expression for K is a function of the angle α and to determine K, α must be measured, since it it not uniquely defined. It was found in this case to be approximately equal to $\pi/4$ radians. Equation 12 can now be expressed as:

$$\Delta r = -0.0708 \cdot \Delta x \qquad \text{(Eq. 13)}$$

Equation 13 expresses the fact that, in the model, the radii of both semicircular lobes diminish in a linear manner when an extension is applied to the sensor. This equation is experimentally verified

by applying an extension Δx to a 0.6 mm core diameter glass optical fibre in the figure-of-eight configuration and simultaneously measuring Δr with electronic callipers. The results were recorded and a linear regression line of Δr as a function of Δx was obtained. Over a variation of 150 mm for Δx, the regression and the actual measurements have a Pearson correlation coefficient of 0.997, which indicates a high degree of correlation and therefore strong linear behaviour within this region.

The regression equation provides an experimental value of the sensitivity coefficient $\Delta r/\Delta x$ of -0.0815, indicating reasonably good agreement with the model (0.0708). The theoretical and experimental values of the sensitivity coefficient differ by 13.1% of the larger value. This result is encouraging given the relative simplicity of the model.

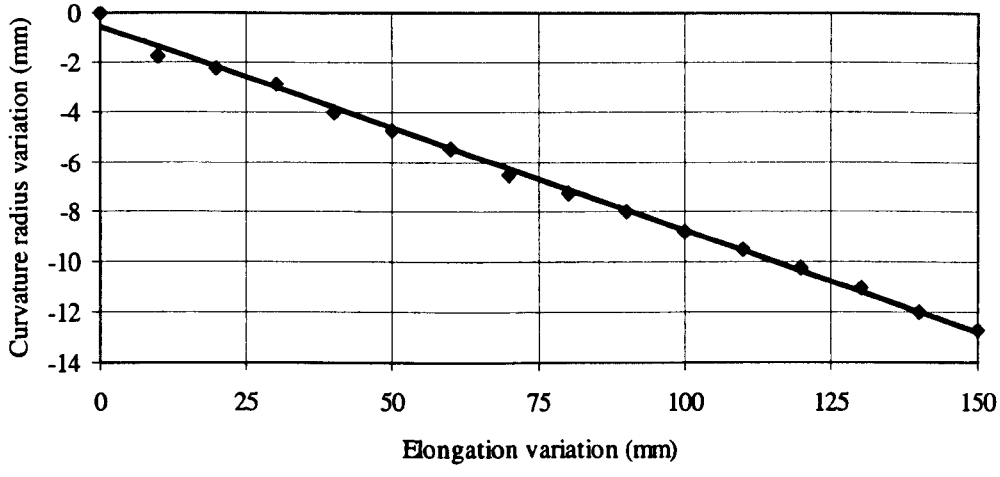

Figure 5: Curvature radius variation Δr in the figure-of-eight coil as a function of elongation variation Δx

In the present state of the model, the most probable source of error is the value of K, which has a strong dependence on α. A plot of the relative error E between the theoretical and experimental values of $\Delta r/\Delta x$ for different values of α is given below. This plot reveals that a 0% percent error between the model and the measured value of $\Delta r/\Delta x$ is obtained when α reaches 60°, which is unrealistic: this value of α was not observed in the figure-of-eight coil. In other words, the relative error is not minimised by more careful measurement of α, which indicates a more fundamental discrepancy between the model and the reality. This discrepancy is likely to lie in the geometrical simplification that the model imposes on the real system.

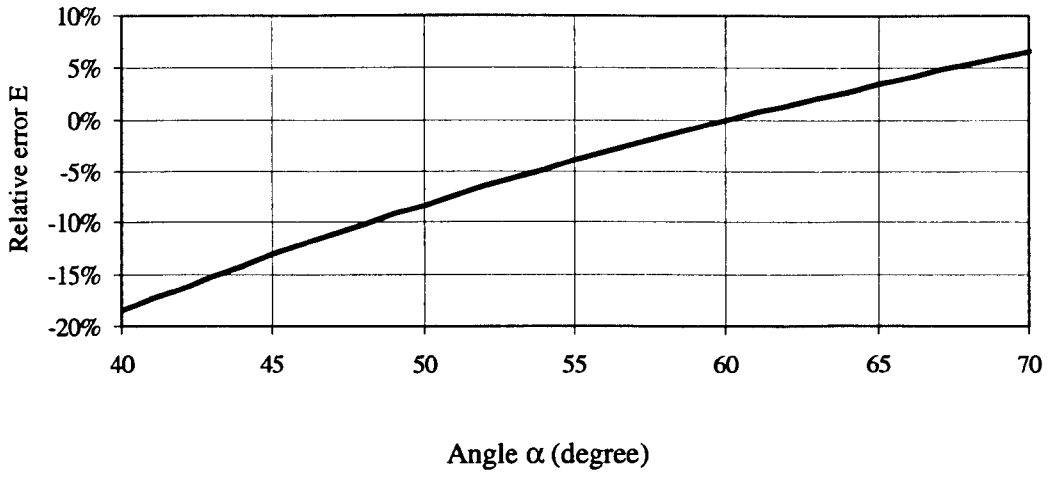

Figure 6: Relative error E for different values of the angle α

9.4. Experimental study of the figure-of-eight response

9.4.1. Experimental conditions

To apply elongation to the figure-of-eighth sensor, an automated test bench consisting of a computer-controlled tensiometer was developed. It was found that the neoprene belt did not alter the figure-of-eight response, so the figure-of-eight was directly mounted on the tensiometer's jaws.

To limit microscopic "stictional" effects (where the figure-of-eight expands or contracts in a series of jumps, rather than by smooth movement), a low viscosity water-based gel was applied at the friction points of the optical fibre.

The tensiometer was programmed to apply elongation to a figure-of-eight coil almost up to its breaking point and then return to its position of origin. All figure-of-eight responses were acquired using the same optical source and receiver, the same fibre input optical power and the same offset and magnification settings. The optical source illuminating the sensor input was an infra red LED (SIEMENS SFH450) of wavelength 950 nm. A matching photodiode (SIEMENS SFH250) was placed at the output of the sensor. The sensor was illuminated by pulsing the IR LED source and the optical power received at the sensor output during elongation was converted in synchrony with the source to an electrical signal via analogue signal conditioning electronics and then sent to the control computer for digitisation and acquisition. Further details of the complete tensiometric, optronic and acquisition systems are given in chapter 10.

To obtain a large dynamic response in the figure-of-eight sensor, we actively contributed to maximising the MBLE. Study of the MBLE in section 4.3 reveals that this can be achieved in practice by overfilling the fibre at the optical launch point in order to exaggerate the number of cladding modes, thereby encouraging the fibre to lose these higher order modes during bending.

The numerical aperture (NA) of the fibre defines a threshold in the range of fibre acceptance angles. This threshold separates the mode population into two groups: core guided modes and cladding guided modes. A small NA will favour the latter and consequently an earlier MBLE response to bending whereas a large NA will concentrate the optical power loss around small radii of curvature.

From a geometrical optics viewpoint, selecting a short wavelength source can also help to maximise the MBLE by increasing the maximum possible number of modes. The chosen wavelength must however conform with the optical fibre transmission spectrum. A source wavelength of 950 nm was, for most optical fibres tested in this research, a good compromise between ease of light transmission and number of modes.

9.4.2. Candidate optical fibres

The choice of a particular type of optical fibre was dictated by a combination of mechanical constraints such as flexibility and elasticity during the sensor elongation. Despite their extremely good flexibility, plastic core optical fibres were discarded because of their almost total lack of elasticity, resulting in unacceptable hysteresis. Silica optical fibres were preferred because of their better elasticity, but their small flexibility made them fragile upon extensive bending. Bare fibres, rather than cabled fibres, were employed to maximise the flexibility and therefore provide the sensor with a long elongation range. Step index fibres were used because they show a larger attenuation by MBLE than graded index fibres [Ghatak & Thyagarajan (1998)]. The choice of fibre diameter is a compromise between two factors: to limit fragility during bending, it was found that fibre outside diameters had to be more than 500 μm. On the other hand, optical fibre of outside diameter larger than 1 mm were discarded because of their rigidity. The range of diameters enclosed between 500 and 1000 μm corresponds to multimode fibres. The number of candidate optical fibres came down to three. They are presented in table 1 below.

Model	NA	Core		Cladding		Coating	
		Diameter (mm)	Material	Diameter (mm)	Material	Diameter (mm)	Material
F-MBC	0.37	400	Pure silica	430	Polymer	730	Tefzel
Unlabelled	0.35	390	silica	420	Polymer	710	Polymer
F-MBD	0.37	600	Pure silica	630	Polymer	7040	Tefzel

Table 1: Candidate optical fibres employed during the experimental study of figure-of-eight coil response. The "unlabelled" optical fibre was termed in this manner as its name and origin could not be traced.

9.4.3. Sensor stretch results

The test bench (described in section 10.4) allowed us to compare the response of the candidate optical fibres under elongation in a figure-of-eight configuration, and the test results are presented and compared in this section. This comparison allowed to determine which of the three fibres had the most adequate behaviour.

The response curve of each fibre in figure 7 below is the average of the response curves during elongation and relaxation. At first sight, we observe that all sensors behave in a similar qualitative manner: the MBLE effect is greater for long elongation (i.e. small radii of curvature) than for short ones. Quantitatively speaking, the unlabelled fibre presents the steepest gradient and the largest response range. The F-MBC fibre reveals similar characteristics to the unlabelled fibre, with a smaller response range and a smaller gradient. Both unlabeled and F-MBC fibres have a core diameter of 400 μm. Finally, the F-MBD fibre, of core diameter 600 μm, has a very small response range and a smaller extension range than the two previous fibres. This is due partially to the stiffness of the F-MBD, which prevented large elongation. Consequently the response range was restricted.

It was found that the unlabelled fibre response curve could safely admit a linear approximation (R^2 = 0.9938) for the elongation range 175 to 200 mm, i.e. within 25 mm prior to the breaking point. The breaking point was reached when the figure-of-eight length became less than 20 mm. In this region, the sensitivity of the unlabelled fibre was 32.5 mV/mm. Given that we used a 12 bit analogue-to-digital converter of input range 5 V, this signifies that the sensor resolution was about 40 μm, which represents 0.16% of the linear region. This percentage indicates that the sensor is well suited for the proposed application.

The F-MBC fibre also admitted a linear approximation within the 25 mm prior to the breaking point, but its correlation coefficient was slightly smaller. (R^2 = 0.9757)

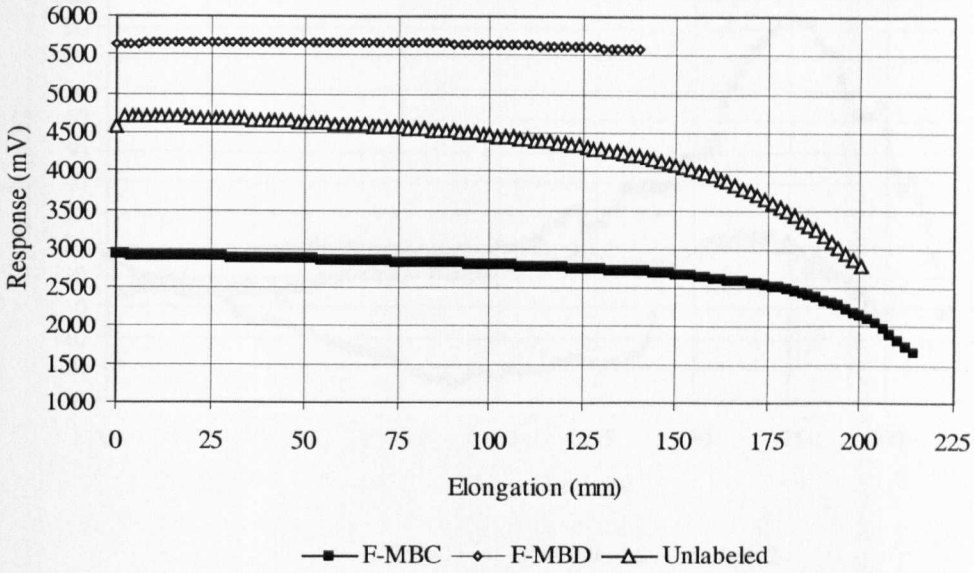

Figure 7 : Comparison of the figure-of-eight coil response to elongation for the candidate fibres.

From the data acquired during the stretching and relaxation elongation tests presented above, the hysteresis curve of each figure-of-eight coil was calculated. The hysteresis curve is defined as the difference in the sensor response between the elongation and relaxation parts of the tests. Figure 8 below presents the hysteresis curve of all three optical fibres. The combined analysis of sensor responses and hysteresis allowed us to conclude which of the three optical fibres was the more adequate.

The unlabeled fibre has the smallest hysteresis curve and its maximum value is 20 mV, which represent 0.5 % of the response magnitude at the same elongation. This fibre was therefore the best candidate since it had a large elongation range, a large response range and a small hysteresis. It is the fibre that was preferred for the construction of the prototype sensor belt.

A positive hysteresis curve is as expected: when the sensor is being compressed back to its original size, its two optical fibres extremities are not completely straight, thus allowing the semicircular portions of the figure-of-eight to be slightly smaller than during expansion. The negative hysteresis curve of the F-MBD is surprising a priori because it suggests that the semicircular lobes are larger during compression than stretching, for an equal elongation. This is actually due to the stiffness of the F-MBD: Because of the stiffness, portions of the fibre outside the figure-of-8 do not bend but remain straight. Consequently, the compression movement has a rotational effect on the figure-of-8, and the effect of this rotation is to delay the radius decrease of both lobes. Hence, for the same elongation, the lobes are larger during compression than stretching and subsequently the hysteresis is negative.

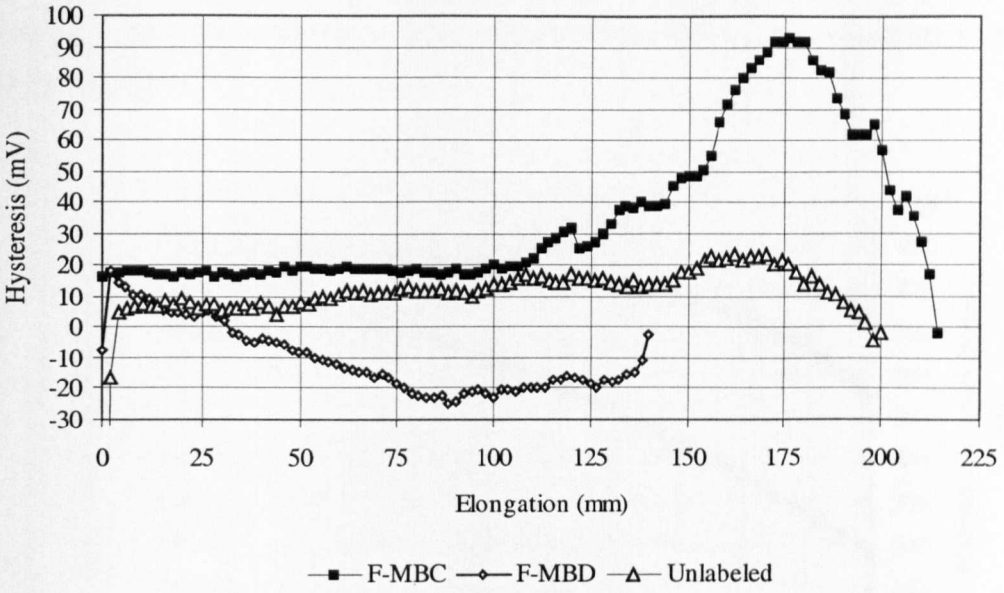

Figure 8: Comparison of the figure-of-eight coil hysteresis during elongation and relaxation for the candidate fibres.

Additionally the unlabelled fibre figure-of-eight coil response was compared to the multiple coil response in order to assess the improvement offered by the figure-of-eight on the multiple coil sensor. The sensors were mounted on the tensiometer in a manner so that a 50 mm elongation range would correspond to the maximum possible elongation before breaking point for both sensors. Figure 9 shows the responses of the multiple coil sensor and the figure-of-eight sensor (made with the unlabeled optical fibre), using the same optical source and receiver, the same fibre input optical power and the same offset and magnification settings.

It was observed that the multiple coil sensor has a relatively large static attenuation (with a small dispersion from the average and a small offset) and a limited elongation range. We attribute these two characteristics to the large number of coils in the sensor configuration, as previously demonstrated in section 7.7.4.1.1, figure 6: Since each coil receives only a fraction of the elongation applied to the sensor, the power attenuation per coil is small and the product of these modulations remains small too. Having the coils aligned in series will necessarily produce a non-linear sensor response, even if the individual coil response is relatively linear in the first place. This is because the overall response characteristic is the product of the responses of the individual coils, which will have the effect of maximising any deviations from linearity. This effect is noticeable in figure 9 below: the multiple coil sensor response curve exhibits significant non-linearity whereas the figure of eight sensor exhibits a more linear response curve. More importantly perhaps, one may also observe that the newer sensor sensor has a response range approximately 2.5 times greater than the earlier one. Since the elongation applied to the figure-of-eight coil is distributed over a single coil only, the variation in the radius of the coil is maximised and consequently so is the MBLE. As identified earlier, the figure-of-eight response presents a very good linearity within the range of 25 mm prior to breaking point (25 to 50 mm on the figure below). In contrast, the multiple

coil sensor does not present satisfactory linearity (R^2 = 0.9112) in this same range, and the only acceptable linear range is between 49 and 50 mm (R^2 = 0.9837). The sensitivity in this region is only 19.1 mV/mm.

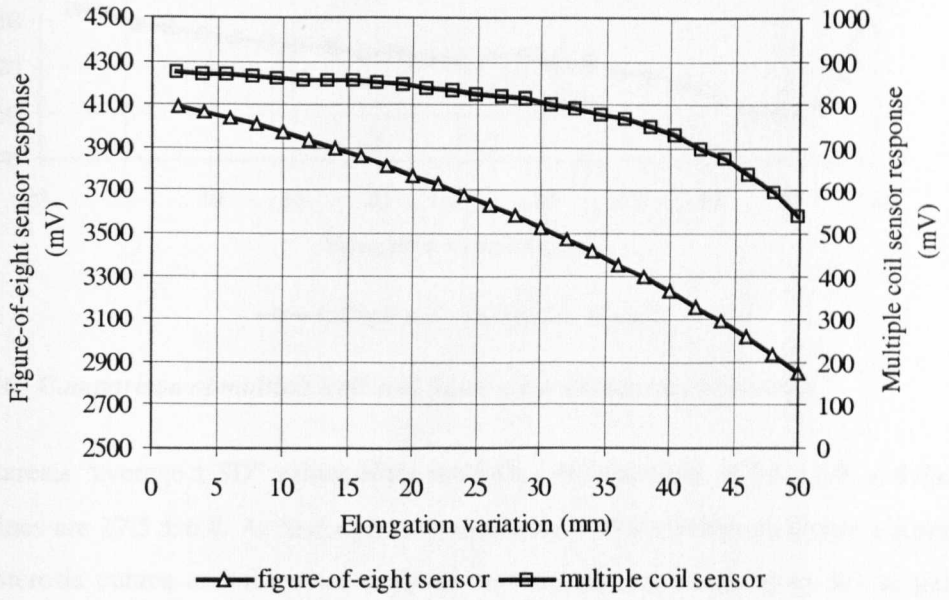

Figure 9:Comparison of multiple coil and figure-of-eight sensors responses Note that both curves have a different amplitude scale

The hysteresis curves of the multiple coil and the figure-of-eight coil sensors were measured and are presented in figure 10 on the next page. In this graph, each curve represents the difference of the signals obtained during the forward and the reverse path. The turning point is situated at the elongation variation of 50 mm.

The multiple coil hysteresis curve is strictly negative. This is surprising since it implies that the optical fibre coils have larger radii during compression than during expansion. This is not perfectly understood but it might be explained by the uneven distribution of elongation among the coils during expansion and compression, as discussed in section 7.7.4.1.

It is also important to note that the figure-of-eight hysteresis is continuously decreasing within the 25 mm towards the breaking point, whereas the multiple coil shows large inflexion 5 mm before the breaking point.

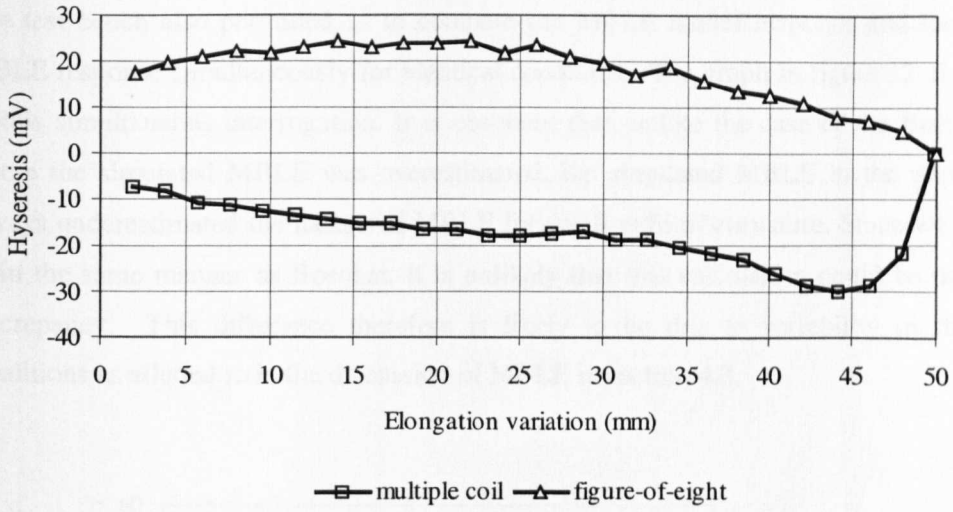

Figure 10: Comparison of multiple coil and figure-of-eight sensors hysteresis

The hysteresis "average ± SD" values of the multiple coil sensor are -17.3 ± 6.9, and the figure-of-eight values are 17.5 ± 6.8. At first sight, the similarity of these values indicates a resemblance in both hysteresis curves and therefore suggests that the figure-of-eight does not actually present significant improvement in the reduction of hysteresis. However, one must remember that the figure-of-eight sensor has a larger response range than the multiple coil. To relate hysteresis and response amplitude, we have represented the relative hysteresis curves of both sensors below in figure 11. The relative hysteresis curve is defined as the hysteresis divided by the average of the sensor response between the upward and downward elongation tests. On this graph, we can see that the multiple coil sensor relative hysteresis shows a maximum of 22% whereas the figure-of-eight remains within the range 0 to 4%.

The "average ± SD" values are $-7.0\% \pm 5.6\%$ for the multiple coil sensor and $2.2\% \pm 0.6\%$ for the figure-of eight. This demonstrates the superiority of the figure-of-eight coil over the multiple coil, in terms of relative hysteresis.

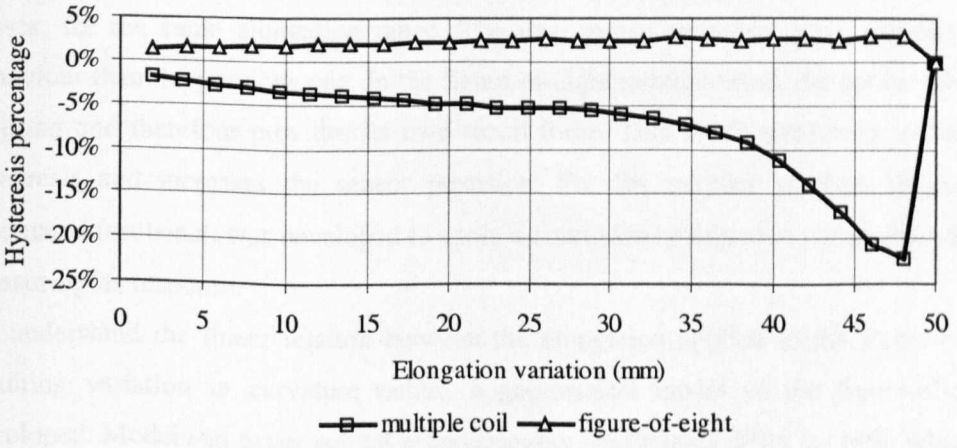

Figure 11: Comparison of multiple coil and figure-of-eight coil sensors relative hysteresis

9.5. Comparison of experimental and theoretical responses

The test bench also permitted us to compare the MBLE model response and the figure-of-eight MBLE response simultaneously for identical conditions. The graph in figure 12 displays the result of this simultaneous interrogation. It is observed that, unlike the case of the Boechat experiment where the simulated MBLE was overestimated, the simulated MBLE in the present experiment always underestimated the measured MBLE for small radii of curvature. Since we have calculated P_r in the same manner as Boechat, it is unlikely that this calculation could be the source of the discrepancy. This difference therefore is likely to be due to variability in the experimental conditions as alluded to in the discussion of MBLE in section 4.3.

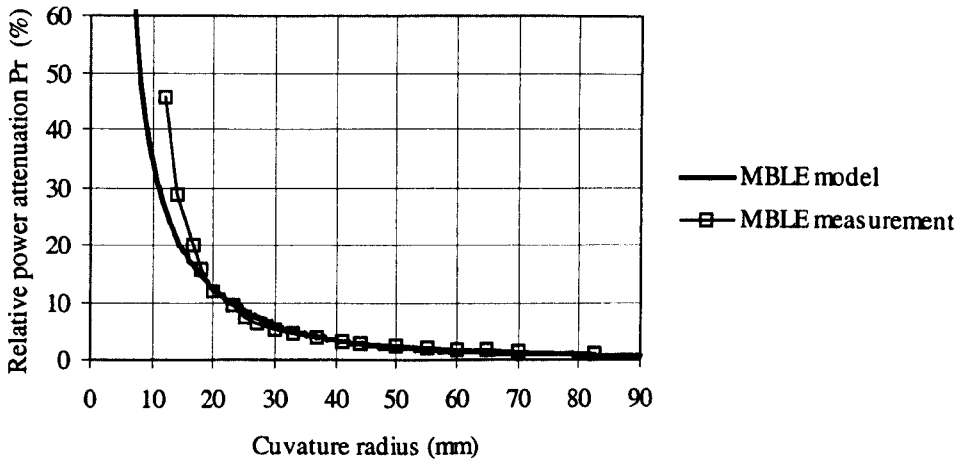

Figure 12: Comparison between MBLE model and figure-of-eight transmission in simulated F-MBC optical fibre (pure silica with NA= 0.37, core radius=0.4 mm)

9.6. Conclusion

The FORP sensor has been redeveloped with success. The new sensor, based on a simpler coil configuration resembling a figure-of-eight knot, has proven very satisfactory in terms of MBLE response amplitude, achieving a response range 2.5 times larger than the response range of the old sensor, for the same elongation range. The new sensor response curve exhibits a more linear behaviour than the previous one. In the figure-of-eight configuration, the optical fibre behaves like a spring and therefore provides its own recoil force. This mechanism helps reducing the sensor hysteresis and increases the sensor precision. For the purpose of these elongation tests, an automated test bench was developed to apply a continuous elongation variation to the sensor while measuring its response.

To understand the linear relation between the elongation applied to the figure-of-eight and the resulting variation in curvature radius, a geometrical model of the figure-of-eight has been developed. Model and experimental proportionality coefficients differ by 14%, which is acceptable given the simplicity of the geometrical model.

The experimental MBLE response of the figure-of-eight under elongation has been compared to a semi-theoretical model of the MBLE given by Boechat for short length large core optical fibre.
It has been observed that the model is in good agreement with experimental data, except for small curvature radii where discrepancies appear, due to the nature of the model itself. It was pointed out that the calculation of MBLE response is largely depending on experimental conditions and therefore no purely theoretical model could make an accurate forecast. Despite these discrepancies, both model and experimental results can be approximated as linear responses for small elongation variation, as in normal quite breathing.

Chapter 10 FORP ACQUISITION SYSTEM DEVELOPMENT

10.1. Introduction

From an instrumentation point of view, the FORP is a system that measures the optical power emerging from an optical fibre-based elongation sensor. The previous version of the FORP has essentially been used as a proof of principle and achieved an accuracy commensurate with the limited requirements of ventilatory monitoring.

It was relatively basic in construction and left considerable room for improvement, in terms of measurement precision. Three factors contributed to the limited performance of the system. First, the sensor had a small linear sensitivity over a small elongation range. Second, the sensor illumination was relatively low. Thirdly, the photodetection system did not permit sufficient amplification to compensate the effects of the second factor, resulting in a small signal.
The ADC resolution was 10 bits and therefore offered a theoretical signal-to-noise ratio of 60 dB. In practice, the photodetction system produced a signal with a dynamic range of only 37 dB. See figure in section 7.7.4.1.2 on page 118.
At the time, the FORP did not offer the capability to reliably and accurately monitor cardiac activity, although it had been observed on occasion.

In this research program, it was decided to design a higher quality instrument that could be regarded as a step towards a professional medical system. The development was motivated by the commercial interest that was revealed by the previous version during visits to various hospitals. As a result, the FORP system has been completely re-designed. The sensor re-development was the object of chapter 9. The new acquisition system is presented in this chapter.

10.2. First prototype design

A first prototype was built with a view to replacing the FORP measurement instrument created by Raza because the latest version had become unusable due to missing parts. The new prototype would also serve as a basis for a more elaborate second prototype FORP measurement instrument.

The first prototype consisted of an emitter and a receiver circuit. The task of the emitter circuit was to continuously illuminate the sensor input. The optical source was an IR-LED (Siemens SFH 450V). The receiver circuit had the task of converting the optical power from the sensor output into a voltage signal that could be easily recorded on an XT plotter. The circuit consisted of a transimpedance amplifier followed by an inverter amplifier and an offset subtractor. See figure 1 for circuit diagram. The system only provided a single one recording channel.

The observation of a cardiac signal at thoracic level was achieved with this first prototype by mounting the figure-of-eight sensor directly on the patient's skin. The resulting signal was of extremely small amplitude, even with maximum amplification using a 15 MΩ feedback resistor. Such large amplification invariably implied large noise level, which hindered a more detailed analysis of the cardiac signal. This system would surely have benefited from the addition of balancing resistors on the positive input of each Op. Amp. to avoid picking up ground noise. The XT plotter came with built-in low pass filters that reduced most of the noise observable at the receiver output.

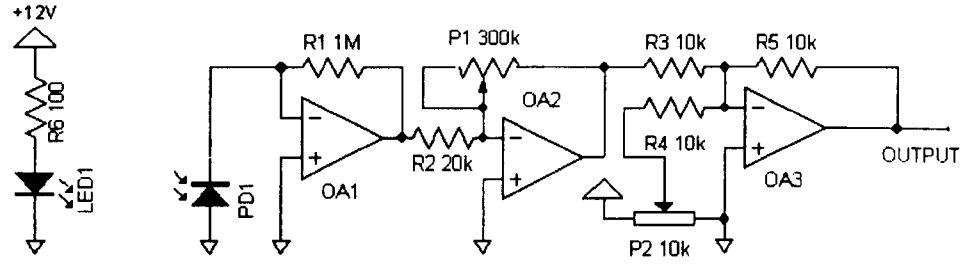

Figure 1: Circuit diagram of the first prototype FORP acquisition system. All Op. Amps are basic UA741. The optical source LED1 is an IR-LED (SFH250V from Siemens) and the optical receiver PD1 is a photodiode (SFH450V from Siemens). Potentiometer P1 acts on the current to voltage ratio, and potentiometer P2 acts as a voltage source for offset subtraction from the FORP signal. The sensor, located between LED1 and PD1 is not represented here.

10.3. Second prototype design

A second prototype was designed because numerous improvements were required for the study of cardiac signals. The motivation for such improvements and their practical implementations are now presented.

The new system, like the earlier one by Raza, had to be capable of digital signal storage for off line analysis and processing. This was achieved with the use of a PC-based architecture: digitised signals were stored on the PC hard drive.

The new system was to provide increased digitisation resolution compared to the earlier prototype. This was achieved by using a 12 bits ADC (max SNR = 72 dB), instead of the original 10 bit model (max SNR = 60 dB).

The noise level of the analogue conditioning circuits was to be decreased. Although adequate low pass filters were to be added to the new system, diminishing the noise level partially consisted of increasing the optical power into the sensor.

Different optical sources have been envisaged and the IR-LED previously used in the first prototype appeared as the best choice because of its small size (fitting on a PCB) and large numerical aperture (compared to a laser). Pulsing the IR-LED permitted it to reach an instantaneous optical power 10 times larger than in continuous mode. In practice, the IR-LED was driven with a strong current (1.5 A) for duration inferior to 10 µs, every 500 µs. The duty cycle was therefore less or equal to 2%, ensuring sufficient time for the diode to cool down after each pulse. Because the IR-LED was not turned on continuously, such a procedure implied that the optical receiver and the signal acquisition system were synchronised with the optical source at every moment. This strict condition was respected by using a common reference clock. The clock would generate a pulse to trigger the IR-LEDs driver and the acquisition system would respond by acquiring a sample and storing it in the PC's RAM.

Another requirement in the design of the second prototype was the simultaneous acquisition of multiple analogue signals. A minimum of two channels was required for the thoracic and abdominal signals but additionally another two channels were added for the recording of signals from the spirometer and the ECG machines.

Finally, the new system had to be designed for a low cost, preferably using the components that were available in the laboratory at this time.

10.3.1. Architecture of the second prototype

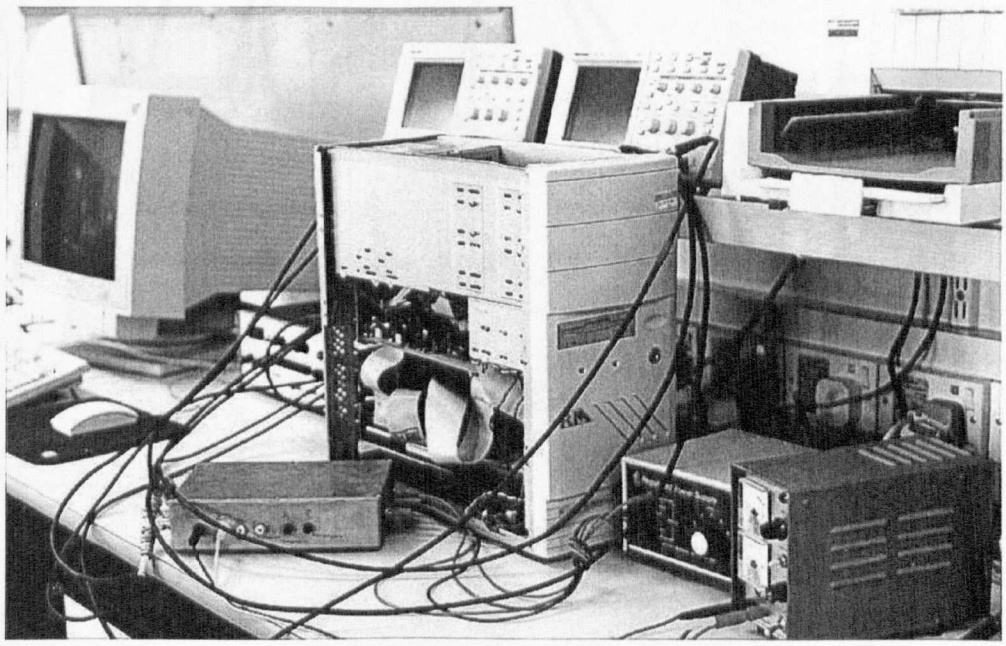

Figure 2: General view of the new PC-based FORP prototype. Note that most elements of the system are contained within the PC, except the optronic system (flat grey box located in front of the PC) and the two power supplies (located on the right side of the PC).

A PC served as the backbone of the new FORP system. The majority of the system elements were purpose-built because the condition dictated by system wide synchronisation required that optical pulse generation and signal acquisition were intrinsically linked to each other, in a manner that only tailor-made hardware and software would permit.

The new FORP prototype is basically made up of 6 elements:

- An IBM compatible personal computer with an Intel 486 DX4 processor, clocked at 100 MHz.
- An 8 bit ISA card containing two digital Input Output (IO) circuits (Intel 8255) capable of generating CPU interruption and a single programmable timer (Intel 8254) that served as a common reference clock for the system. This card is hereafter termed the IO/timer card.
- A pulse conditioner card for the amplification, inversion and filtering of the pulses controlling optical signal emission and signal acquisition.
- An optronic system for pulsing the IR-LED and conditioning the sensor output signals
- An analogue to digital converter (ADC) card to acquire the conditioned signals.
- A C++ written control application for the operation of all the elements mentioned above and the management of the PC resources

The Optronic system, the ADC card, the pulse conditioner and the control application were developed and prototyped in the laboratory. The IO/timer card and the PC were the only manufactured elements. All these elements are now presented in detail.

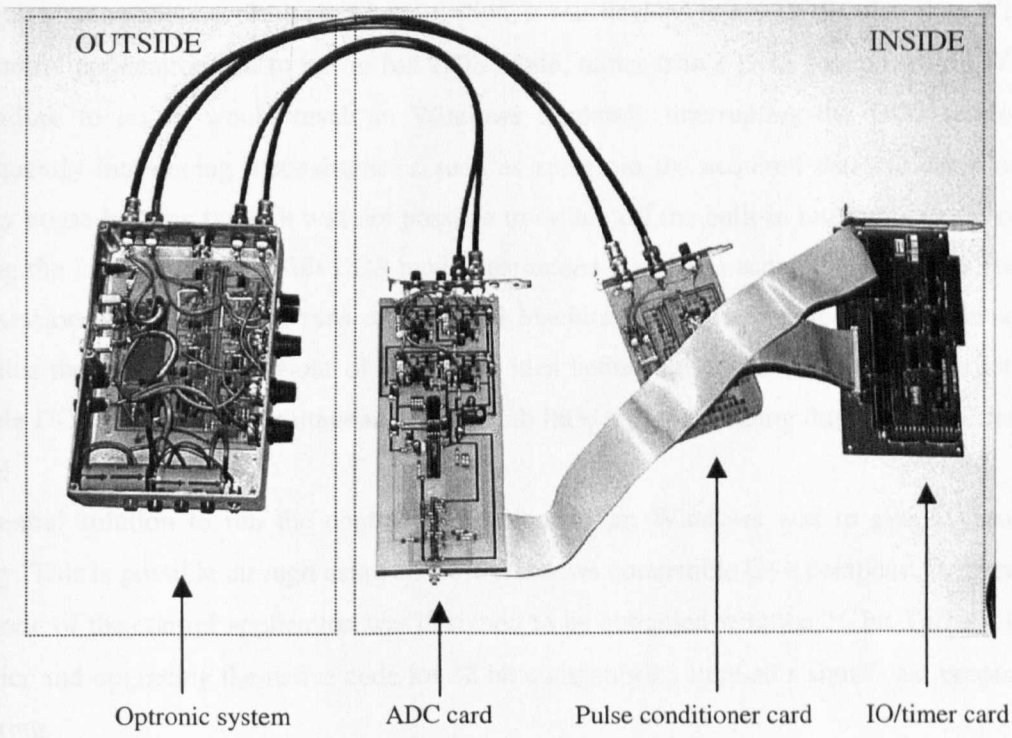

Figure 3: General architecture of the new FORP prototype

10.3.2. Personal Computer and control application

It was decided to use a PC as the backbone of the second prototype. This is because a PC environment furnishes ready to use services such as volatile and non-volatile memories, user interface hardware (mouse, keyboard, screen), casing, power supply, networking, a huge choice of compatible peripherals, and last but not least, low level programmability.

Additionally, with the fast increase in computational power and memory size achieved in the recent years, PC prices have considerably diminished, making even systems only a few years old easily affordable. After a suitable PC was chosen, all bus-connected peripherals were taken out, except the graphics card. The hard drive was reformatted and the machine was reinstalled with Microsoft Windows 95. The interest in using this operating system is that it comes with a second operating system: MS DOS 7.0. DOS permits easier access to the PC hardware resources (and in particular hardware interruption handling) than Windows. The latter requires the development of specific "protected mode" drivers for hardware access. Additionally, DOS is almost a truly real-time operating system in the sense that it dedicates CPU time to only one application at a time, whereas Windows allows sharing CPU time in between different applications by allocating CPU time-slices according to a fairly complex algorithm governed by 4 priority classes: "critical", "real-time", "user" and "background".

For equivalent results i.e. clean signal acquisition, it was decided to use DOS rather than Windows. The control application had to run in full DOS mode, rather than a DOS session within Windows and failure to do so would result in Windows randomly interrupting the DOS session (and consequently introducing inconsistencies such as spikes in the acquired data) to carry out high priority house keeping tasks. It was not possible to switch off the built-in interruptions. In contrast, running the DOS program in full DOS mode guaranteed clean data acquisitions. This is because a DOS session within Windows runs on a Virtual Machine (VM), i.e. a software environment that simulates the hardware behaviour of a PC. The idea behind the use of VM was to permit to run multiple DOS application simultaneously[1] but with little regard to timing during real PC hardware control.
A potential solution to run the control application within Windows was to give it "real-time" priority. This is possible through using a 32 bit Windows compatible C++ compiler. Originally, the C++ code of the control application was designed to be compiled with the 16 bit Turbo C++ 3.01 compiler and upgrading the native code for 32 bit compatibility implied a significant proportion of re-writing.

---

1 Each application running in a VM appears to run on its own individual computer. This allows applications that were not designed for multitasking to run concurrently with other applications.

It was therefore decided to keep the DOS version of the control application, since it performed satisfactorily. The control application was given the name "SYNACQ", which stands for synchronous acquisition. The final version of control application code can be found in appendix 1.

The tasks of the control application consisted of:

- Interfacing with the user, which includes asking how many acquisition channels were required and for how long, and where to save the data files.
- Initialising the hardware resources, which includes programming the 8254 timer and the 8255 IO chips, setting up the interruption vector.
- Allocating the required memory
- Performing the signal acquisition (by simply doing nothing but waiting for the ISR to be called)
- Transferring data to the hard drive

The acquisition ISR itself consisted of:

- Masking interruptions (to prevent nested interrupts from disrupting the acquisition ISR)
- Reading data from the acquisition card via the IO chips
- Buffering data to conventional memory[2]
- Unmasking interruptions

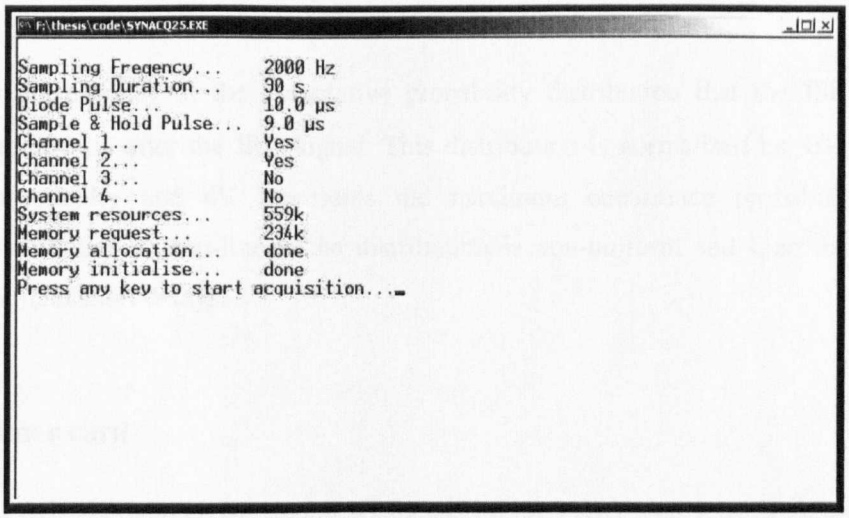

Figure 4: Screenshot of "SYNACQ", the control application user interface

2 During acquisition, the conventional memory (640 kB) was almost empty, apart from the upper 100 kBytes. About 540 kBytes were left unused, which permitted an acquisition duration of 35 seconds at a sampling frequency of 2 kHz on all 4 channels. In practice, such a duration was sufficient and not all 4 channels were always required.

Directly writing to the hard drive from the ISR did not yield very good results because random access to the hard drive would provoke spikes to appear in the ground line of the PC power supply, therefore corrupting the analogue signals to sampled. It was noted that there was a consistent delay of duration in the range 24 to 36 μs between the IRQ signal (the interruption is triggered on the falling edge of the IRQ signal) and the ISR activation, as can be seen in figure 5. Even if a delay between the two events was expected, the duration of this delay was surprisingly long. This delay implied that the FORP signal should be buffered, prior to ISR activation. This is achieved with a sample and hold (SH) amplifier. See section 10.3.5 for details on the SH amplifier.

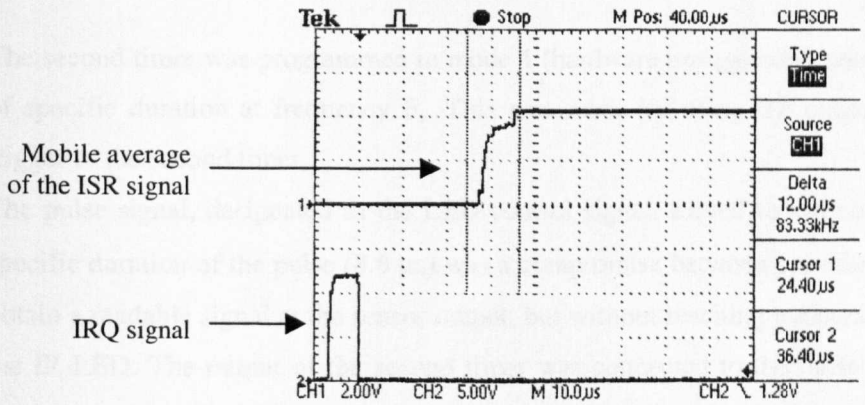

Figure 5: A screenshot of the oscilloscope measuring the delay between the IRQ signal and the ISR activation. The ISR signal was generated by the first 8255 chip on the IO/timer card. The top trace on the oscilloscope screen represents the mobile average of the ISR signal in the time window 24 to 36 μs.

This actually corresponds to the cumulative probability distribution that the ISR might occur between 24 and 36 μs after the IRQ signal. This distribution is normalised i.e. 0V represents no occurrence probability and 4V represents the maximum occurrence probability. Since the cumulative distribution is non-linear, the distribution is non-uniform and long delays are more likely to occur than short delays.

10.3.3. IO/timer card

The IO/timer card generated all the digital IO signals for the FORP system to communicate with its elements situated outside the PC case. Instructions to operate the acquisition card were issued by the ISR, which was called upon request from the IO/timer card by generating hardware interrupt signals to the CPU. Finally, the IO/timer card was also used for flagging particular software events, such as ISR activation.

The main advantage of this card was to offer a ready to use platform for simple interfacing of external systems with the PC. Operating the IO/timer card in real time was achieved by the control

application. The IO/timer card supported the very popular industry standard 8254 and 8255 chipset. The operation of both chips with respect to the functioning of the system is now described.

With its three independent and programmable 16 bit timers, the 8254 solves one of the most common problems in computer programming, i.e. the generation of accurate time delays under software control. Instead of setting up time loops in software, the first timer was used as the common reference clock of the FORP system. It was programmed in mode 0 (prescaler counter) to generate a square wave signal running at a frequency of 2 kHz. This actually was the sampling frequency F_s of the system. The 8254 master clock was the ISA bus clock (4 MHz) of the PC.

The second timer was programmed in mode 1 (hardware retriggerable one-shot) to produce a pulse of specific duration at frequency F_s. This was made by using the output of the first timer as a trigger to the second timer.
The pulse signal, designated as the LED control signal, served to switch on the 2 IR-LEDs. The specific duration of the pulse (8.6 μs) was a compromise between providing a long enough time to obtain a readable signal at the sensor output, but without reaching temperatures that would damage the IR-LED. The output of the second timer was connected to the pulse conditioning card, itself connected to the IR-LED drivers and finally the IR-LEDs. Given that the pulse signal has a duty cycle of 1.7%, the IR-LEDs were off more than 98% of the time, thus ensuring that they were not damaged by overheating.

The third timer was also programmed in mode 1 (hardware retriggerable one-shot) and produced a pulse signal of frequency F_s = 2 kHz and duration 7.6 μs in synchrony with the second timer. The resulting pulse signal, designated as the SH control signal serves two purposes.
First, its falling edge triggered the sample-and-hold (SH) amplifier (in the signal conditioning circuit of the optronic system) so that both sensor output signals would be momentarily frozen by the SH amplifier. This was so that the ADC could perform a conversion on a stable voltage. The duration of the SH control signal (7.6 μs) was chosen so that the IR-LEDs currents had reached a plateau permitting stable optical signal measurements.
Secondly, the SH pulse served to ask the CPU to call for the Interrupt Sub Routine (ISR) containing the necessary instruction that operated the acquisition card. Such a request was made by sending the SH control signal to the strobe input (pin C4) of the 8255 circuit in mode 1. In return, the 8255 generated an interrupt flag (pin C3) that was received by the ISA bus of the PC and passed on to the interrupt priority decoder. The selected interrupt request (IRQ) was line number 5. IRQ 5 was the first free high priority hardware interruption line that was available from the 8 bit ISA bus. The IRQ was triggered on the falling edge of the SH pulse, like the SH amplifier.

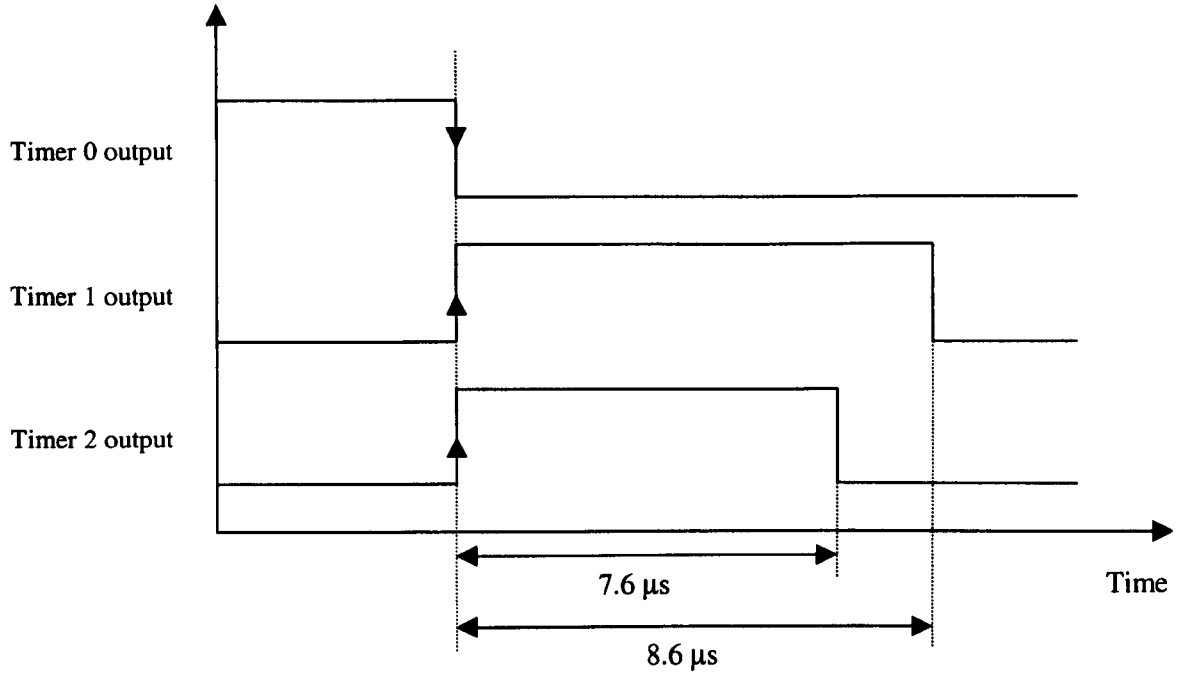

Figure 6: Chronogram of the 8254 timer outputs.

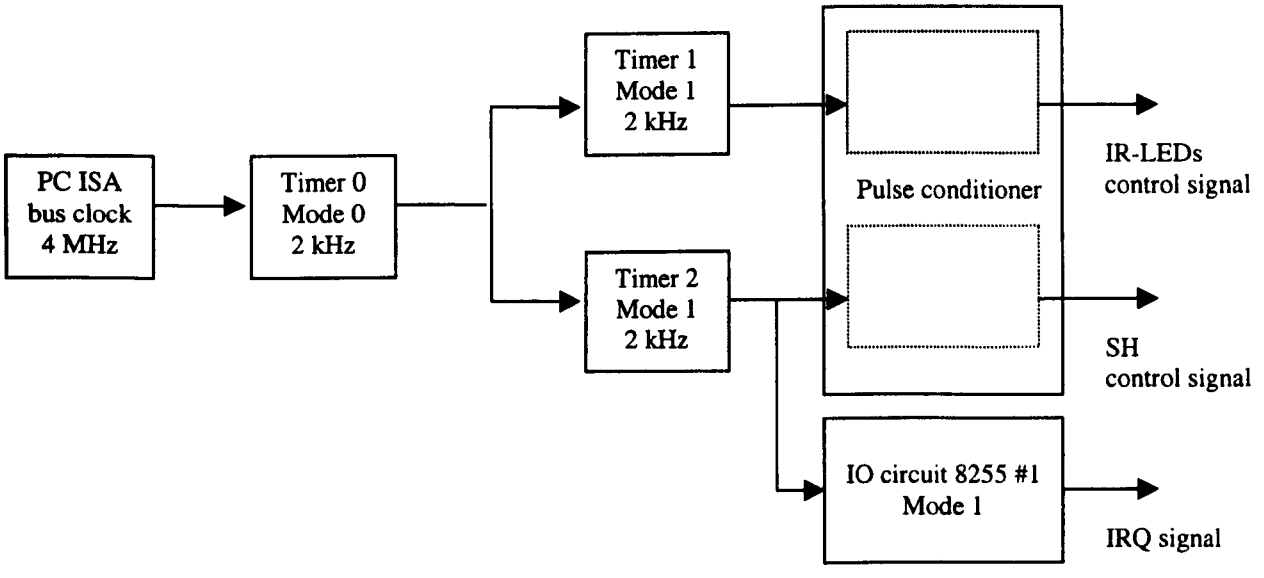

Figure 7: Structure of the timer

The 8255 chip is a general purpose programmable digital IO device which provides 24 IO pins that may be individually programmed in two groups of 12 IO pins and used in 3 major modes of operation. The IO/timer card came with two 8255 chips and both of them were fully employed for various tasks. As reviewed above, the first 8255 was programmed in mode 1 (strobed IO) which permitted the use of port C to generate an IRQ to the ISA bus, on demand of the first timer of the 8254. Port B was configured as an output and was freely used to flag different events in the control application, such as the beginning of the ISR for example[3].

The second 8255 chip was programmed in mode 0 (basic IO mode) and served to communicate with the acquisition card. The wiring of the 8255 to the acquisition card consisted of control lines (ports A & C) and data lines (port B). 3 lines of Port A (A0 A1 A2) were dedicated to the control of the analogue multiplexer located prior to the ADC on the acquisition card, and 3 other lines of Port A (A3 A4 A5) served as control signals to the ADC. The 8 lines on Port B were entirely wired to the data lines of the parallel output on the ADC. Because the ADC had a 12 bit resolution, the 8 most significant bits were accessed by reading port B and the remaining 4 bit were accessed by reading again pins 0 to 3 of port B. Port C was used as an input to monitor the status of the ADC.

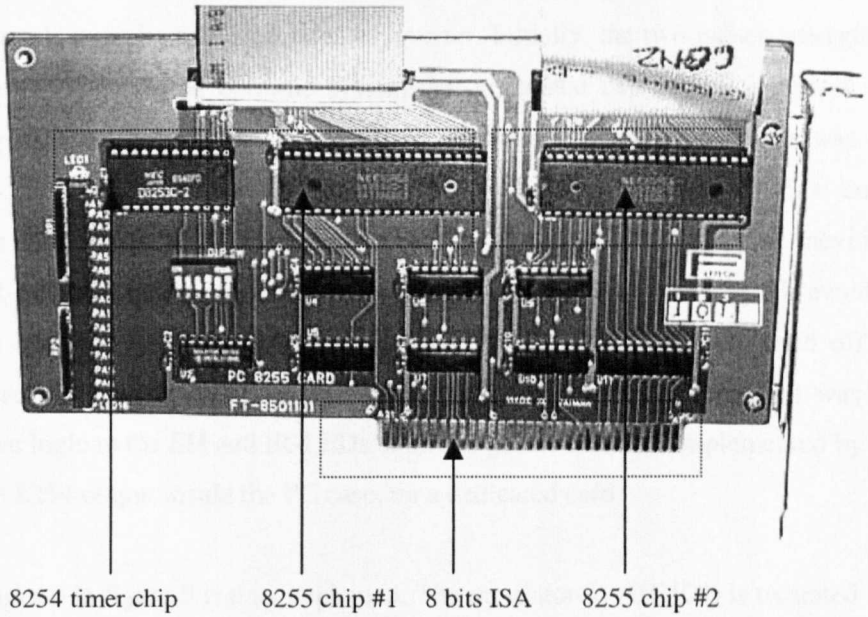

Figure 8: The IO/timer card layout. The left ribbon cable is connected to the pulse conditioner and the right one to the acquisition card.

3 This technique of flagging software events (especially the ISR) proved to be a very successful help during debugging of the control application. It permitted the measurement of the timing, a critical parameter in a real time system, of virtually any instruction in the original code.

10.3.4. Pulse conditioner card and IR-LEDs driver

The pulse conditioner card and IR-LEDs driver are two elements of the same functional section aimed at producing the optical input to the sensor and triggering the sample and hold (SH) amplifier to "freeze" the sensor output signal before acquisition.

The pulse conditioner card (located into the PC) contains an identical pair of circuits that produce the SH and IR-LEDs control signals by conditioning the output signals of timers 1 & 2 of the 8254 timer chip. The conditioning operations consist of low pass filtering[4] and inverting. The SH and IR-LEDs control signal are sent to the optronic system via coaxial cables (BNC1 and BNC2) where they are dispatched then to the IR-LED driver and the SH amplifier.

One may wonder why the pulse conditioning operations were separated from the optronic box. In fact the isolation of the pulse conditioner from the IR-LEDs driver is strategic; this is now explained. The pulses emerging from timers 1 and 2 of the 8254 use negative logic, which implies that a "True" is manifested by a ground potential and a false by a positive potential. On the contrary, the SH circuit and the IR-LEDs driver have a positive logic and compatibility between the two logic protocols must be provided with an inverter. Initially, the two pulses emerging from the 8254 were directly sent to the optronic system that contained two logical inverters. Despite its apparently favourable logistics, the disadvantage of this configuration however was that, if the computer was accidentally switched off before the optronic system, the IR-LEDs control pulse would become a ground potential, therefore switching on the IR-LEDs for a time inevitably far too long for the IR-LEDs to survive the dramatic heat that they were experiencing. To avoid liquefying both IR-LEDs and their PSU each time the computer was accidentally switched off before the optronic system, or each time the IR-LEDs control signal was disconnected, it was decided to provide positive logic to the SH and IR-LEDs control signals. This was implemented by placing the inverters at the 8254 output inside the PC case, on a dedicated card.

The circuit diagram in figure 9 is now explained. The transistor T_1 (BC109) is mounted in common emitter mode, which provides the inverting effect. A simple first order low pass filter is formed by adding a capacitance between the base and the emitter of T_1. During variations of the voltage at the input of the conditioner, the current emerging from R_1 momentarily sinks into the capacitor C1 and the base current is therefore low pass filtered. The resistor above T_1 is branched to the computer's 5V PSU output. The collector signal of T1 is the IR-LED control signal. Similarly, the collector signal of T4 is the SH control signal.

---

4 Low pass filtering was required because the IR-LEDs control signal had the responsibility of modulating the power going through the IR-LEDs and, given that currents up to 1.5 Amps are crossing each IR-LED, it was important that the control signal did not contain any noise, spikes or sharp corners that might have inadequately injected too many charges into the transistors responsible for driving the IR-LEDs.

Both control signals are then passed on to the optronic system (via coaxial cables BNC1 & BNC2) where they are separated. The SH control pulse is sent to an amplifier (not represented in figure 9) and then to the SH amplifier. The IR-LED control signal branches to the IR-LED driver, capable of injecting large currents into the IR-LEDs. At the input of the IR-LEDs drivers are transistors T_2 and T_5, which act as positive amplifiers. The collector signals of T_2 and T_5 are sent to T_3 and T_6, also acting as positive amplifiers but with a larger current driving capability, thanks to the use of Darlington pairs. The voltage across each IR-LED is about 2.3V when the current through them is pumped up to 1.5 Amps. It must be noted that such a current is more than 10 times larger than the maximum acceptable current in continuous mode (130 mA). The IR-LEDs were operated on the edge of their maximum capacities and for this reason, they sometimes displayed slightly different characteristics. To limit the difference in optical power between the two IR-LEDs, a population of 10 IR-LEDs was studied and the two most similar behaviours were matched. It was observed that a V_{CX2} higher than 12 Volts would occasion premature breaking down after a few hours of operation. A "standard" V_{CX2} value of 10.4V was adopted for normal operation.

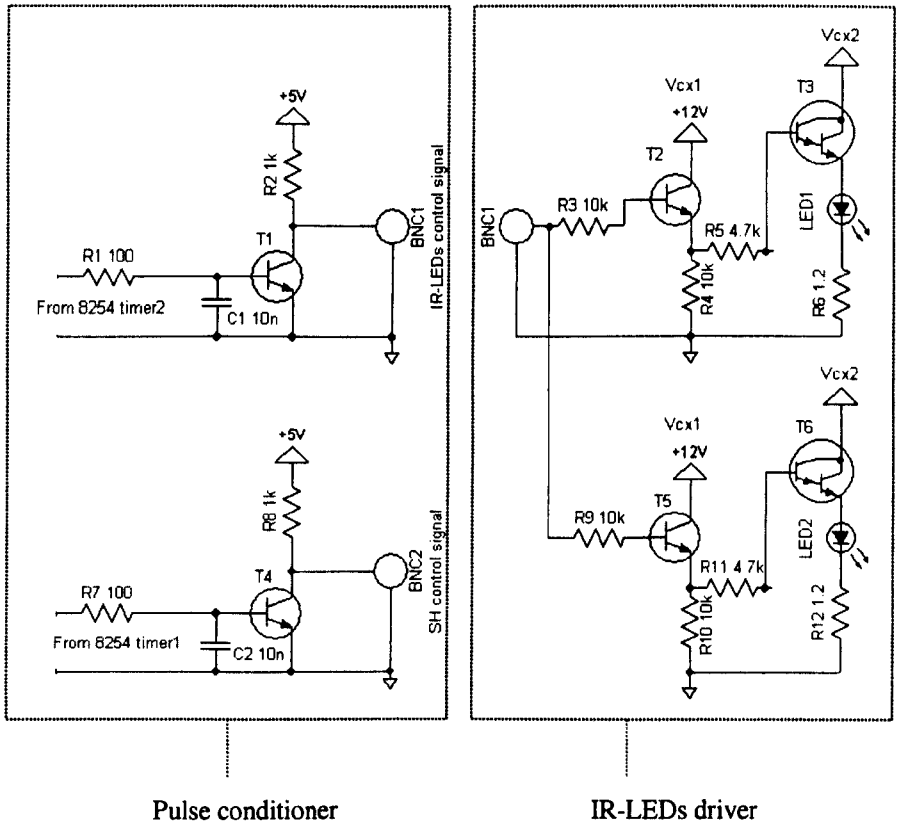

Figure 9: Circuit diagram of the pulse conditioner and IR-LEDs driver. T1, T2, T4 and T5 are the NPN transistor BC109 for general purpose audio application. Its slow rise time rounds off the edge of the control pulses, hence providing a low pass effect that prevents from brutalising the IR-LEDs with large current variations. T3 and T6 are NPN Darlington pair MPSA13 to handle large currents. LED1 and LED2 are IR-LEDs SFH250V.

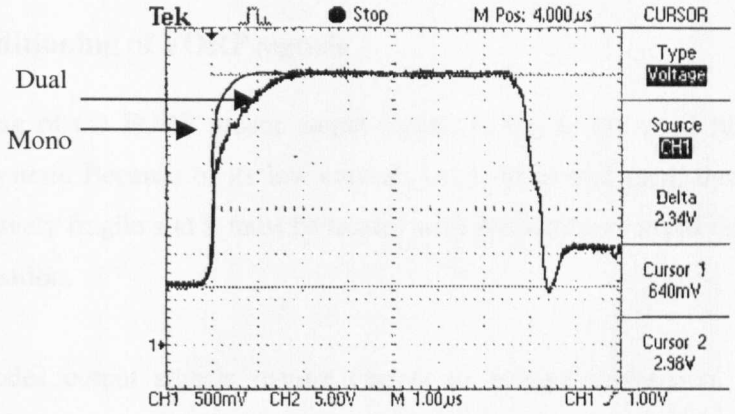

Figure 10: Screenshot of the oscilloscope measurement of the voltage across one IR-LED in the absence of the second IR-LED (The trace is called MONO) and in the presence of the second IR-LED in parallel (The trace is called DUAL). The value of Vcx2 is 10.4V and the plateau voltage value is 2.34V.The effect of the second IR-LED is to soften the rising edge of the voltage pulse. This is because, with two IR-LEDs, twice as much current is pumped from the PSU and consequently, it takes longer to obtain a stable current. The slow reaction of T1, T2, and T5 (See figure 9) also contributes to the soft knee in the curve.

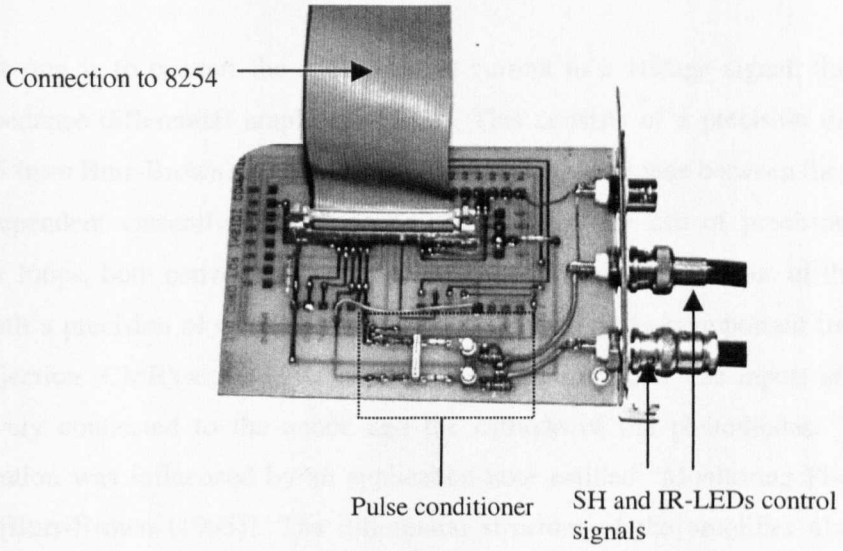

Figure 11: The pulse conditioner card. The ribbon cable carries pulses from the pulse generator (the 8254 chip) and the conditioned pulses are wired to the optronic system by BNC connections.

10.3.5. Conditioning of FORP signals

The processing of the FORP sensor output signals is key to the good functioning of the whole acquisition system. Because of its low current, large offset and small dynamic range, the FORP signal is relatively fragile and it must be treated with precaution to avoid distortion and consequent loss of information.

The photodiodes output signals require current to voltage conversion, offset subtraction and amplification before being digitised. The digitisation process occurs in two steps, the signal is first held by the sample and hold (SH) circuit and the momentarily frozen signal is low pass filtered and then fed to the input of the ADC. More specifically, the conditioning operations required to make the FORP signal ready for digitisation are discussed here.

Conditioning operations are carried out by circuits based on precision instrumentation operational amplifiers. Two identical conditioners are required for the treatments of the two FORP signals. The circuit diagram is shown below in figure 12.

The first step is to convert the sensor output current to a voltage signal, this is done with the transimpedance differential amplifier (TIDA). This consists of a precision differential amplifier (INA105 from Burr-Brown) of unity gain that calculates difference between the signals received by two independent current/voltage converters. Thanks to the use of precision resistors in their feedback loops, both converters have a gain precisely equal to the value of the resistor, i.e. 3.01 kV/A with a precision of 0.1%. Having exactly identical gains is important for efficient common mode rejection (CMR) at the input of the differential amplifier. The inputs of the converters are respectively connected to the anode and the cathode of the photodiodes. The choice of this configuration was influenced by an application note entitled "Monitoring Photodiodes with Op. Amps" [Burr-Brown (1995)]. The differential structure of the amplifier also ensures that the photodiodes are not referred to ground, while limiting the gain of each current/voltage converter, which permits them to have a larger bandwidth. Short, shielded cables serve the wiring of both photodiodes. The TIDA provides an amplification gain precisely equal to 6.02 kV/A.

The negative voltage output of the TIDA is branched to a negative adder so as to subtract the offset of the FORP signal. Consequently, the output of the adder is a positive signal whose offset has been minimised. This signal is passed to a positive amplifier of variable gain, in the range 1 to 20000. The total maximum gain of each signal conditioner is 120 MV/A (or 0.12 V/nA) which makes it possible to measure photodiode current in the order of a few tenth of nanoamps.

The reason for placing the adder before the amplifier is to more accurately minimise the offset. If the amplifier was in front of the adder, then the offset of the raw FORP signal would be amplified and it is more than likely that the signal would be saturated because of its enormous offset.

The amplifier output then branches to the sample and hold amplifier. This very circuit is a key factor for the signal quality of the whole system since it is extremely prone to degrading the signal. Particular attention has been given to the implementation of this critical component, in particular on the PCB by isolating and shielding signal tracks from control and power tracks.

The SH amplifier output is fed to a follower which serves as an impedance adapter prior to sending the signal to the acquisition card via a 50 Ω coaxial cable.

Pulsing the IR-LEDs involves currents of about 1.5 Amp, which is a major source of disturbance for the rest of the circuit, considering that signals smaller than a few nA are generated by the photodiode on the same printed circuit board.

The circuit being its own source of noise, considerable care has been given to noise limitation. All analog parts of the system are strongly shielded and the optronic system case is earthed. On the PCB, the shield is provided by large pads of ground tracks, as can be seen in figure 13, and especially in the IR-LEDs driver section. These pads have the effect of developing a large capacitance between the surrounded signal tracks and the ground, therefore absorbing all current spikes. Particular attention was given to the cabling. Inside the optronic system, wiring was made with shielded cables of various diameters. BNC connectors and coaxial cables provided connection to the optronic system. Finally, the PSU was filtered using two large capacitors of 4700 μF each, connected across the power lines (± 12V) and the PSU ground. The same shielding techniques were employed on the acquisition card PCB and additionally, individual capacitors were added between the power lines of each integrated circuit.

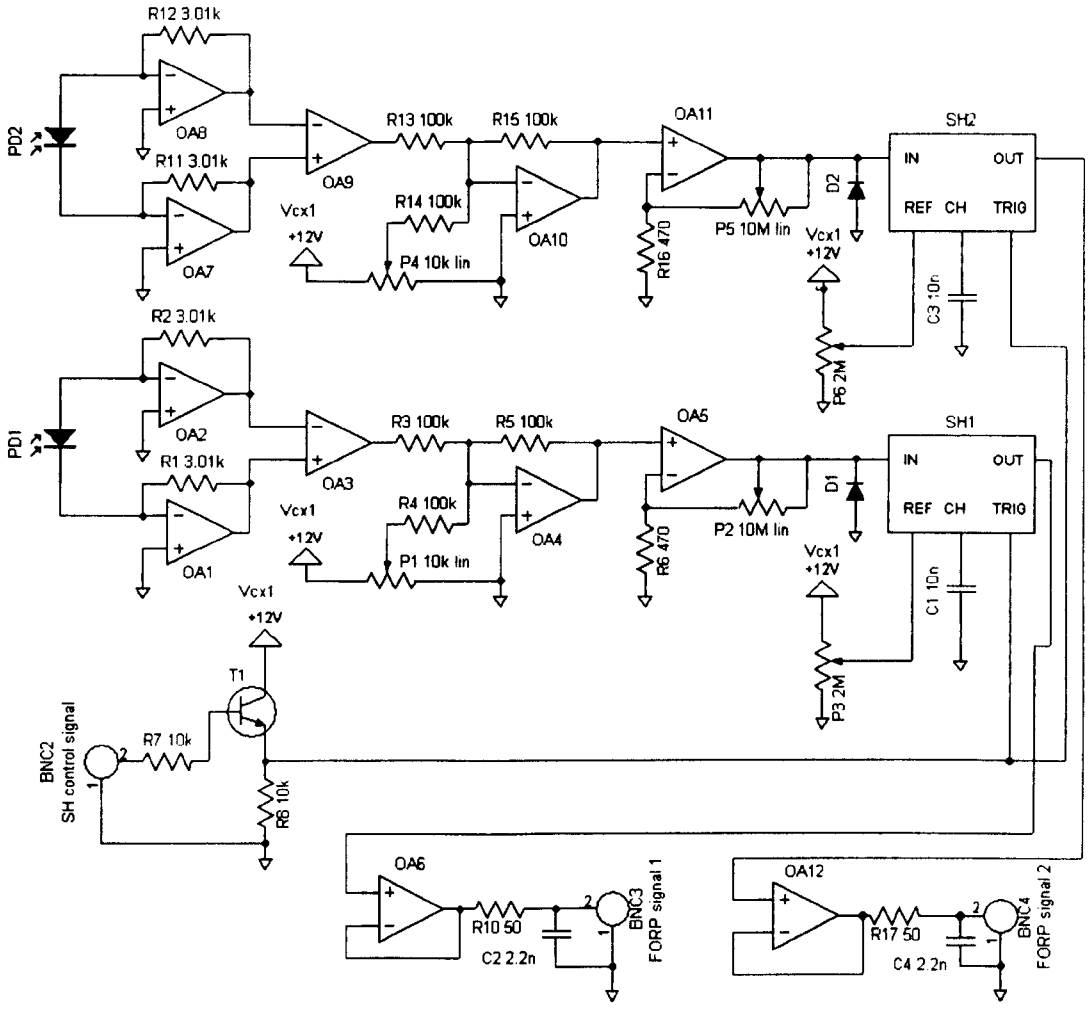

Figure 12: The FORP signal conditioner circuit. The design of the second channel is completely identical to the first one and shares the same SH control signal. All Op. Amps are OPA2111 except OA3 and OA9 which are INA105. PD1 and PD2 are photodiodes SFH250V. SH1 and SH2 are sample & hold amplifiers LF398A. T1 is a NPN transistor BC109. BNC2 is the SH control signal input. BNC3 and BNC4 are FORP signal outputs to the acquisition card.

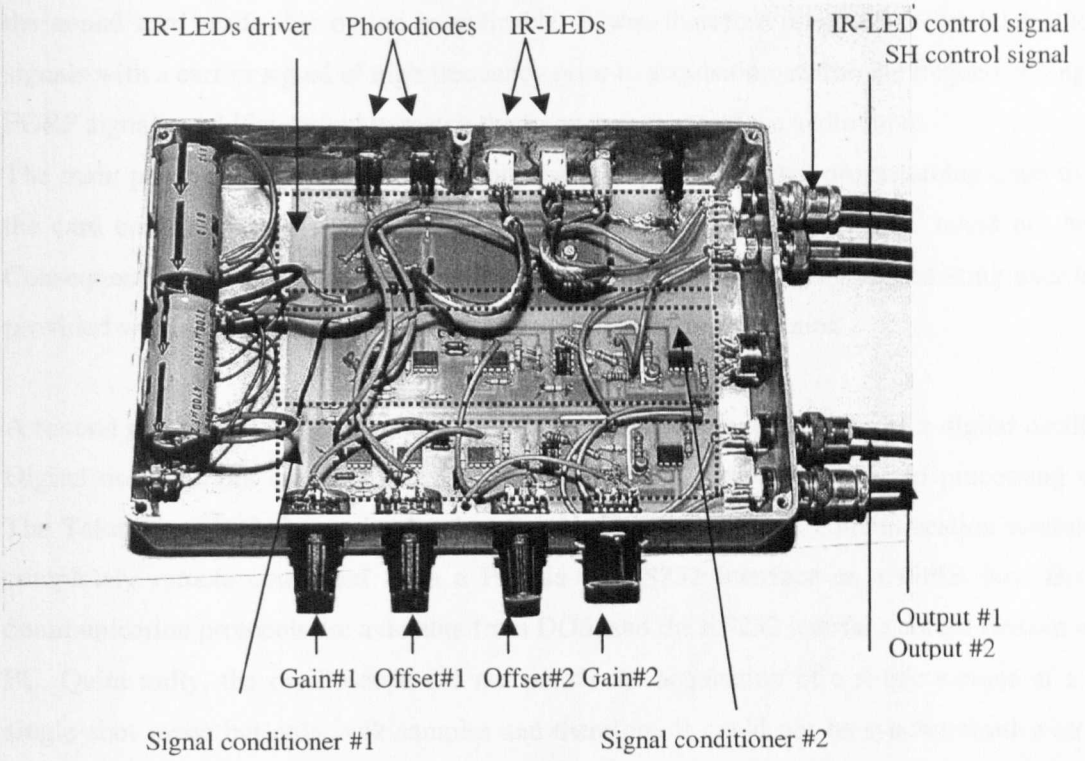

Figure 13: The optronic system, containing the IR-LEDs driver and the two signal conditioners.

In the present configuration, both aspects of the optronic system, the IR-LEDs driver and the signal conditioners are successful at achieving the tasks they had been designed for. However, if it were to be redesigned, the IR-LEDs driver would be separated from the signal conditioners to improve immunity to EM disturbance created by the pulsed current.

10.3.6. Acquisition card

As an alternative to the use of conventional acquisition systems, it was initially considered to use the sound card of a PC. The advantages of this idea were numerous. First, sound cards have very good quality signal acquisition hardware (16 bit resolution on two independent channels, sampling frequency up to 48 kHz in most cases, samples are transferred to the hard drive in real time). Second, sound cards are so popular that they are now inexpensive and found in almost any PC, and so are the control applications to operate them.

One technical difficulty however in the use of a sound card is that most of the audio signal frequency range is within 20 to 20 kHz and consequently, built-in high-pass analog filters are implemented on the sound card to cut off signals of frequency below 20 Hz.

Given that the respiratory signal measured by the FORP system has the vast majority of its components below 20 Hz, it was envisaged to physically remove the components of the high pass

filters. The use of surface mounted components and the absence of technical drawings concerning the sound card made this option inapplicable. It was therefore proposed to modulate the FORP signals with a carrier signal of high frequency prior to acquisition so that the frequency range of the FORP signal would be shifted to match the frequency range of the audio input.
The main problem with the sound card approach however was that programming code to operate the card could not be written because the card low level control operation[5] could not be found. Consequently, the card could not be used outside the range of the already existing user interface provided within Windows and, for this reason, the idea was abandoned.

A second potential solution for the acquisition system came with the use of a digital oscilloscope. Digital oscilloscopes are, in essence, analog-to-digital conversion and signal processing systems. The Tektronix TDS210 permitted, when fitted with the TDS2CM communication module, to be completely remote controlled from a PC via an RS232 interface or a GPIB bus. Both these communication protocols are available from DOS, and the RS232 interface comes built-in with any PC. Quite sadly, the oscilloscope did not permit the acquisition of a single sample at a time in single shot mode but only bulk samples and therefore, it could not be synchronised with the IR-LEDs. This solution had to be discarded.

The third potential solution for the acquisition system was to use a dedicated ISA acquisition card called ADC42 from Blue Chip Technology. The card offered a 12 bit ADC (MAX172 from Maxim) and 8 independent input channels. The card also supported IRQ generation for the 8 bit ISA bus and digital IO with the rest of the world via a single 8255 chip. Additionally, it was straightforward to operate as most of the control code could fit in a few lines of C++. More importantly perhaps, the implementation diagrams and the components list were fully available. The card gave very good results during test and presented itself as the best available solution. Unfortunately, it was rendered unusable as a result of a polarity inversion in the power feed lines during a test. Virtually all components on the card were burnt or suspected to malfunction, including the on-board firmware contained in a PIC microcontroller. Because Blue Chip Technology would not re-issue the firmware as a spare part, it was decided to abandon the ADC42.

Two other acquisition card were tested (ADDA12 from FlyTech, and PC30 from Eagle) but they both appeared unsatisfactory in terms of noise level in the digitised data.

5 Low level control designates the programming code required to operate the card without relying on a pre-existing interface in Windows.

Finally, it was decided to build a prototype acquisition card based on the ADC42 card. More specifically, it took advantage of the IO/timer card to interface the ADC chip with the PC. Such procedure guaranteed rapid prototyping by eliminating the need to develop specific ISA compatible hardware. The IO/timer card would provide ready-made PC bus access and IRQ generation. Figure 19 represents a picture of the prototype.

A strategic choice for the acquisition card was to place the analog multiplexer in front of the ADC so that the 4 analogue inputs could have completely independent ranges, via the use of 4 separate conditioners[6]. Each conditioner was based on a precision single chip dual operational amplifier (OPA2111 from Burr Brown) and permitted to freely modify the gain and the offset of each of the 4 input channels (see figure 14). This configuration eliminates the need for a specific external conditioner per input channel and all input signals could have different ranges. In practice, modification of individual gains and offset was rarely used, because the potentiometers were difficult to access, being located directly on the card.

An extension of this concept was to "auto-condition" each input channel independently so that the acquisition range would always be optimal with respect to the input signal range. This would eliminate the task of manually finding adequate gains and offset settings for all acquisition channels. The implementation diagram of the auto-conditioning system is presented in figure 15. Despite good results obtained during development, it could not justify the prototyping of a second acquisition card.

Control of offset and amplification circuits was achieved with digitally controlled potentiometers, or DCP[7]. The gain and offset values would be determined at the beginning of an acquisition session, according to the statistical analysis of the first few seconds of acquired data. It can be observed in figure 15 that the first DCP is mounted as the front resistance of the negative amplifier. This may be surprising because it means that the gain of the amplifier is then inversely proportional to the DCP resistance. However, it is easily shown that the inverse of the gain is proportional to the input range and it yields that the DCP resistance is directly proportional to the signal input range. If the DCP had been mounted as the feedback resistance of the negative amplifier, then its value would have been inversely proportional to the input signal range. Calculating the inverse of the input voltage range is of no concern when using a digital system with a large floating point precision (like the PC) but it will produce unexpected results when the system only provides limited floating point precision. This phenomenon is usually termed finite word length effect. In fact, the auto-conditioning system was initially developed to be operated by a PIC 16C74A

---

6 On most acquisition card, the analog multiplexer is located prior to the signal conditioner, itself branching to the ADC. The result is that all analog inputs must have the same input range, which is not necessarily the case, especially when acquiring signals originating from different sources.

7 Modulation of gain and offset could also be generated with the use of a Voltage Controlled Amplifier (VCA). However, the vast majority of VCAs have a non-linear gain response to control voltage and DCP were preferred for their inherent good linearity.

microcontroller from Microchip and such a device does not handle floating point numbers at all. In this condition, it is preferable to avoid dividing numbers as the ratio of two integers is not necessarily an integer, resulting occasionally in dramatic precision loss. On the contrary, multiplying two integers always generates an integer and this explains why the DCP was preferred in the frontal position. The circuit would have benefited from the addition of balancing resistors on the grounded inputs of each Op. Amp.

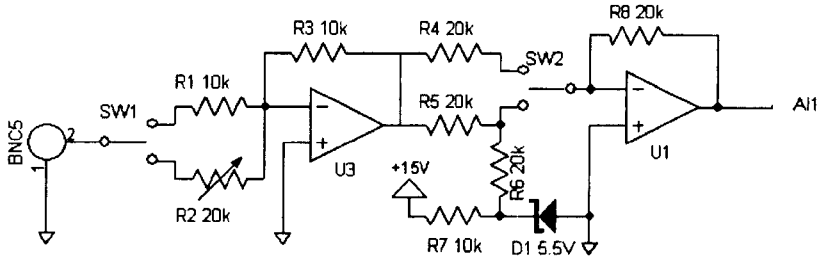

Figure 14: One of the four conditioning circuits for the analogue input signal of the prototype acquisition card. The circuit was design to receive a signal with maximum dynamic range ≤ 10 V, either unipolar or bipolar. All four conditioning circuits are identical. The two amplifiers are OPA2111 precision instrumentation operational amplifiers. All resistances, expect the DCPs are 0.1% precision resistors The first Op. Amp. is mounted as a negative amplifier, and the second op. Amp. is mounted as a negative adder. The negative factors in each transmittance cancel out and V_{out} has the same polarity as V_{in}. The switch SW1 gives a choice of unity gain (input range 5V max) or variable range (input range 0 to 10V max). The switch SW2 is for unipolar/bipolar selection. The question of whether the adder should come second or first in the circuit line-up depends on the characteristics of the input signal. If Vin is characterised by a small dynamic range and a large offset (like the FORP), then it would be preferable to place the adder before the amplifier. If the signal has a small dynamic range and a small offset, then it is preferable to have the amplifier first so that the offset can be extracted more precisely. All power feed lines to the Op. Amps (+15V, -15V and ground) are supplied by the on-board power supply NMA1505.

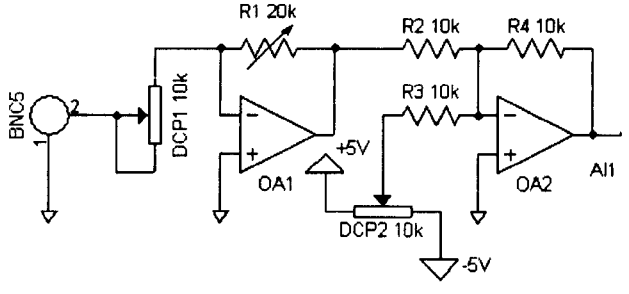

Figure 15: Possible implementation diagram for the auto-conditioning system. This system is based on the manual conditioner. The digitally controlled potentiometers are X9C103 from Xicor. They provide 100 steps of resistance variation. Balancing resistors should have been added on the positive inputs of each Op. Amp.

After the signal conditioner comes a 4 to 1 analogue multiplexer, a voltage follower and finally an ADC chip. Figure 16 represents the corresponding circuit diagram, where AI1 to AI4 are the outputs of the conditioners. The voltage follower provided a small output impedance in order to accommodate the small input impedance of the ADC chip. This was required to ensure that the

ADC input voltage could receive sufficient current during conversion, while not draining too much out of the analogue circuit.

The ADC chip on the prototype acquisition card was a MAX172 from Maxim, as employed on the ADC42 acquisition card. The choice for this 12 bit ADC chip originates from its 8 bit parallel output port that rendered the chip very easy to interface with the 8255, as opposed to ADC chips using a serial output. Also, this chip originally equipped the ADC42 and good results were obtained with it in terms of acquisition quality. (When acquiring the same reference signal, it was established by visual inspection that the ADC 42 always generated the smallest noise level compared to other acquisition card in the lab such as the ADDA12.)
The possession of the ADC42 general circuit diagram permitted the replication of the hardware implementation of the chip onto the prototype acquisition card. The ADC chip and the analog section of the ADC42 used their own low noise stabilised power supply (NMA0515 from Harris) and this solution was also adopted for the prototype.

To test the card, its analog inputs were fed one by one with a triangular wave and the resulting digital signals were enormously amplified with Matlab to study their noise level, linearity and monotonicity.
Generally, the prototype gave very satisfactory results in terms of digitisation quality. The noise level was slightly lower than the ADC42, and for this reason the prototype acquisition card was considered to be a successful replacement.
One area of shade however is the relative slowness of ADC operation, being limited by the speed of the IO chip 8255 to which it was connected. The simplest solution to this problem was to decrease the sampling frequency when all 4 channels were used. The best potential solution would have been to speed up the acquisition card by embedding a microcontroller containing code to operate the ADC autonomously. This would have obviously simplified the communication between the ADC and the 8255 and hence reduce the operation time. Again, this solution was adapted from the ADC42 where on-board firmware was dedicated to operating the ADC chip. The added benefit of this solution is that an embedded microcontroller would have permission to trigger the acquisition process with an external signal more easily. This could have been used to permit the SH control pulse to trigger acquisition in hardware, hence freeing the acquisition card from the dependence of the ISR timing.

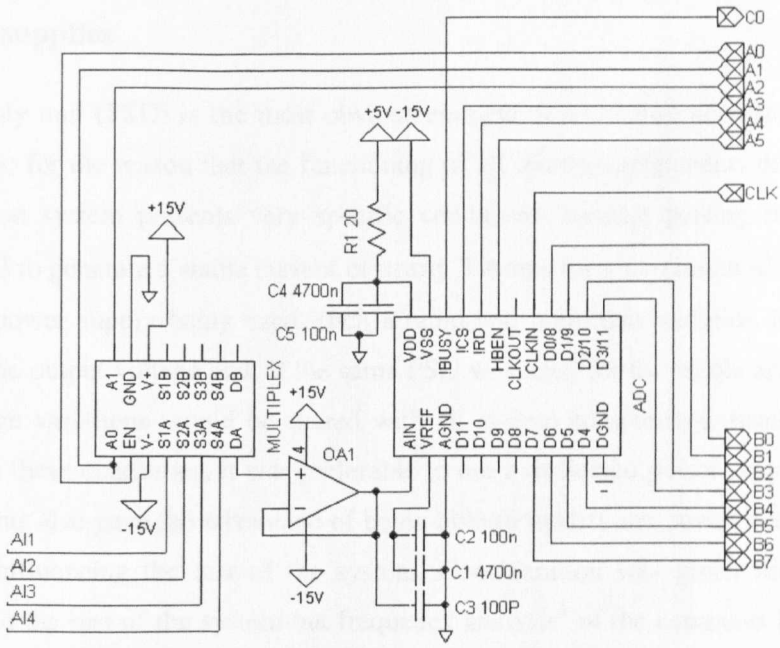

Figure 16: Circuit diagram of the prototype acquisition card, excluding the 4 analog conditioners, which connects with tracks AI1 to AI4. The digital IO ports on the right side refer to the second 8255 IO chip on the IO/timer card, except the clock (CLK) input, which is provided by the 8254. The 4-to-1 analogue multiplexer is a DG409 from Intersil and the ADC is a MAX172 from Maxim. All power feed lines (+15V, -15V and ground) are supplied by the on-board power supply NMA1505, except the ground and the +5V connections to the ADC, which are branched to the PC internal power supply.

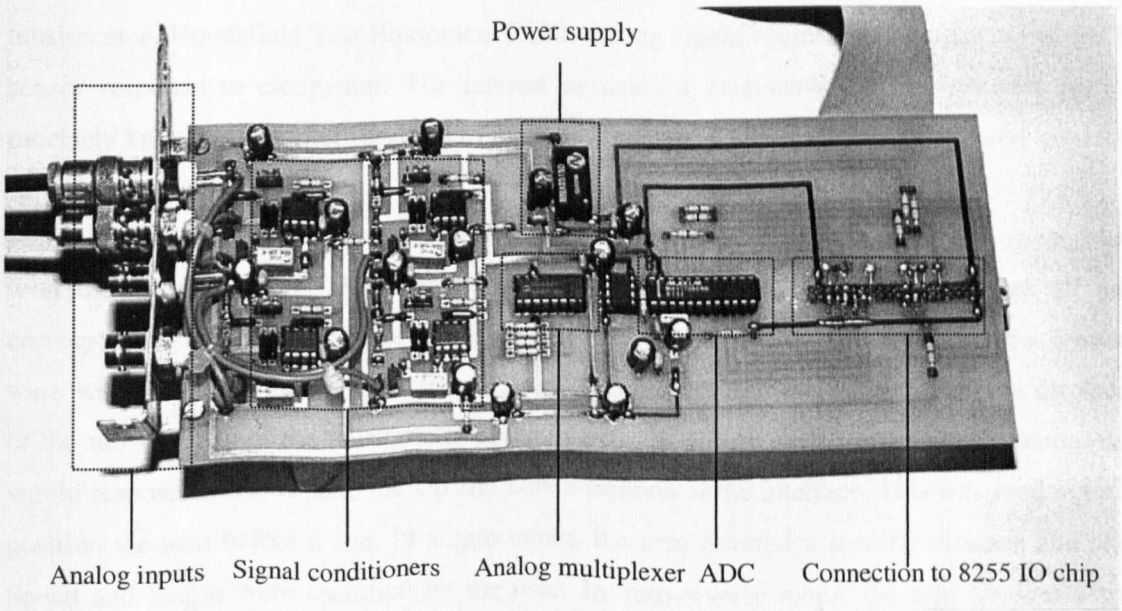

Figure 17: Picture of the ADC42-based prototype acquisition card. Note the double-sided printed circuit board.

10.3.7. Power supplies

The power supply unit (PSU) is the most obvious element of an electronic system, yet often the most problematic for the reason that the functioning of all system components depends on it. The FORP acquisition system presents very specific conditions because pulsing the two IR-LEDs requires the PSU to generate a stable current of nearly 3 Amps for a maximum of 10 μs. Given the old age of the power supply being used, such a rapid and important variation inevitably creates disturbance in the output voltage and, if the same PSU was used for the whole acquisition system, then such voltage variations would be shared with all system components, resulting in potential malfunctions. In these conditions, it was preferable to use a dedicated power supply to power feed the IR-LEDs. This also gave the advantage of being able to modify the power delivered to the IR-LEDs without influencing the rest of the system. Consideration was given for the use of the computer PSU to the rest of the system but frequency analysis[8] of the computer PSU output lines revealed that they were too noisy to be used with analogue circuits. Hard drive operation was recognised as an important source of noise. Unlike digital systems, analogue ones require a very clean PSU output and it was decided to attribute a dedicated 12V bipolar PSU to all analogue sections of the system, except the IR-LEDs.

10.4. FORP sensor test system

The second prototype of the FORP acquisition system was adapted to permit control of a tensiometer (Hounsfield Test Equipment S100) during signal acquisition in order to test the FORP sensor response to elongation. The interest in using a tensiometer is that it could apply very precisely known elongation variation to the FORP sensor, hence precise test data were generated.

A tensiometer control application was written in Visual Basic 5 (VB5) to interface the tensiometer with the acquisition PC. The control application interface can be seen in figure 19 and the corresponding VB5 code is reproduced in appendix 10. The acquisition PC and the tensiometer were wired together by serial communication. The control application permitted the displacement of the moving arm of the tensiometer up and down, in different modes. In manual mode, the arm would respond in real time to the Up and Down buttons of the interface. This was used to precisely position the arm before a test. In single mode, the arm covered a specific distance and stopped. Speed and length were specified by the user. In auto-reverse mode, the arm covered a specific distance, stopped and then came back to the position of origin. Speed, length and directions were specified by the user.

---

8 The Tektronix TDS220 digital oscilloscope, which was commonly used during the development of the acquisition system, served simultaneously as an oscilloscope and an FFT based spectrum analyser, permitting interrogation of both time and frequency domains of a signal.

This mode was used for testing response to elongation and hysteresis of the sensor.
Typical recorded data would consist of a "down up down" return run to stretch and relax the sensor. The acquisition control application generated a file containing the sensor response during testing. Later processing in Matlab permitted the generation of the hysteresis signal.

The acquisition control application was called on the request of the tensiometer control application. An ideal solution would have been to augment the C++ code of the acquisition control application to include management of the tensiometer, rather than developing an entirely new application in VB5. However, Visual Basic renders data transmission through serial interface very easy and therefore guarantees a accelerated development time compared to C++ because serial interface handling comes built-in with VB5 and does not necessitate any further programming.

The instructions sent to remotely control the tensiometer consisted of alphanumerical strings. The strings were sent to the tensiometer in order to activate built-in functions to displace the mobile arm, read the mobile arm position, read the force sensor, lock the front control panel, and other maintenance tasks.

In order to obtain a true sensor response versus displacement, it was necessary to start the tensiometer and the data acquisition precisely at the same time, which required that both systems communicated with each other in real time in order to wait for each other to be ready, regardless of whoever got ready first. Because the two applications could not share the same memory space in the PC[9], it seemed a priori problematic to synchronise them. Failure to do so would have resulted in a response curve out of phase with the elongation applied across the sensor.

The matter of synchronising the two applications was solved by employing previously unused input and output bits of the first 8255 chip (on the timer IO/card) as hardware flags. More precisely, one application would set its own flag and wait for the flag of the other application to be set. When both flags were set, the two applications were both ready to start and proceeded accordingly and in synchrony. The bits employed in this manipulation were bits remaining from IO ports that had already been configured for other tasks. Modification of the 8255 external wiring was involved in the manipulation. Writing/reading to/from the ISA bus from within VB was made possible with the use of an external free DLL called vbio.dll developed by Zealsoftstudio (www.zealsoftstudio.com). Because the tensiometer control application was a Visual Basic application, it had to be run with Windows and consequently, the acquisition application was executed in a DOS window rather than in full DOS mode. At the standard sampling frequency of 2 kHz, this introduced artefacts in the data (see section 10.3.2) and consequently, the standard sampling frequency had to be reduced.

---

9 Windows uses RAM in protected mode, which does not allow separate applications to access the same memory space. This is in order to protect the integrity of the volatile data handled by each application.

Because artefacts in the digitised data took the form of spikes, they could be easily removed with a moving standard deviation filter[10].

The new sampling frequency was left as a variable, which gave the possibility of obtaining a specific spatial sampling frequency, if desired. For example, 1 sample per mm regardless of the arm's speed.

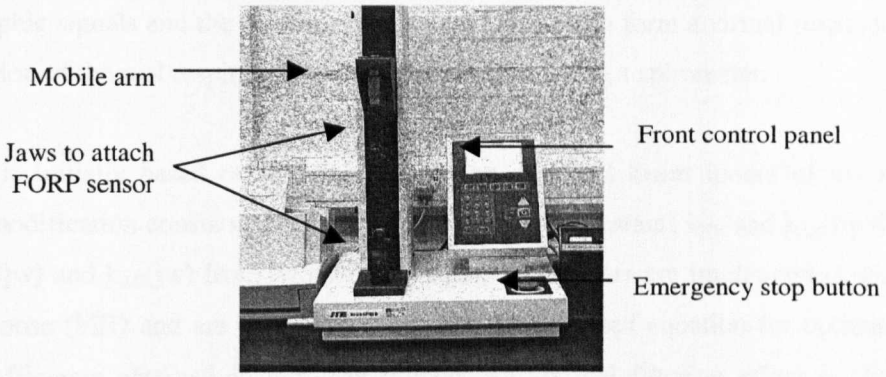

Figure 18: The digital tensiometer S100 from Hounsfield Test Equipment.

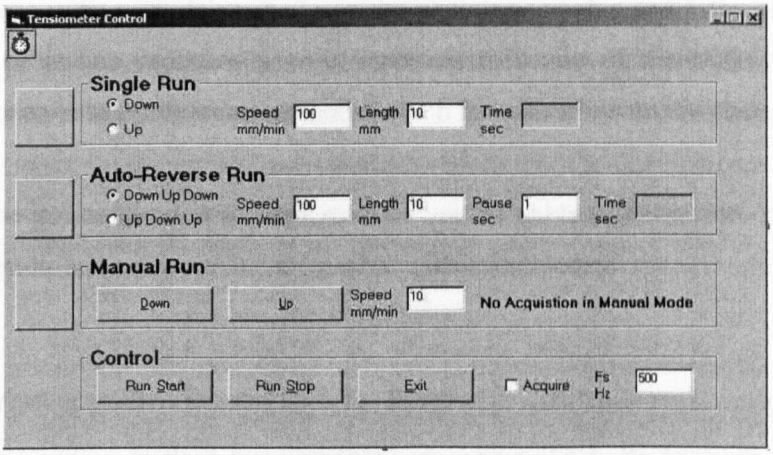

Figure 19: The user interface of the tensiometer control application. It permitted control of the direction, the speed and the distance of the vertical arm displacement. When the run had been rehearsed correctly, the user could tick the "acquire" selection box and the acquisition system was started simultaneously with the arm displacement.

The development of the tensiometer control application and its integration with the acquisition control application were considered successful and permitted to test extensively the stretch response of different optical fibres, as previously detailed in section 9.4.3 and 9.5.

---

10 A moving standard deviation filter is a non-linear filter that calculates the local average and standard deviation of a set of values. Values can be rejected on the basis of their magnitude difference with the local moving average, relative to the local standard deviation. Spikes are easily deleted because they provide large differences from the average.

Chapter 11 NEW CALIBRATION METHOD FOR RESPIRATORY MONITORING WITH THE FORP

11.1. Introduction

This chapter introduces a new calibration method for respiratory monitoring with the FORP, based on the Wiener optimal filter. Two such filters are applied to the thoracic and abdominal plethysmographic signals and the resulting signals are summed to form a virtual respiratory signal, i.e. an estimation of the real respiratory volume, as measured with a spirometer.

This method is initially based on the classical Konno & Mead linear model of the respiratory system. The modification consists of replacing the calibration constants k_{TH} and k_{AB} by the transfer functions $k_{TH}(jw)$ and $k_{AB}(jw)$ from two separate filters. The filters are implemented with a Finite Impulse Response (FIR) and are calibrated using the Wiener-Hopf equation for optimal filtering. The filter coefficients obtained in this way produce an optimal filtering effect in the sense of minimising the mean square error (MSE) between the virtual and the real respiratory signals.

By precisely compensating the delays between the plethysmographic and spirometric signals, this new calibration procedure produces a very important reduction of the MSE compared to the traditional regression based calibration methods (which have no effect on the phase of each signal).

To understand the points that led to implementing such a calibration method, it is important to understand the results and effects of the regressive calibration method.

11.2. The regressive calibration methods for linear and quadratic respiratory models

The regressive calibration method was presented in section 7.6.3. Its principle is to obtain the calibration constants of the respiratory model by minimising the MSE between the virtual and real respiratory signals. This calibration procedure can be applied to both linear and quadratic respiratory models.

When applied to the same group of subjects, it was previously reported that the regressive calibration of the quadratic respiratory model was the method leading to the smallest MSE, hence the best method for accurate respiratory volume estimation [Raza et al. (1998)]. However, the average MSE obtained from the quadratic model was only a fraction smaller than the average MSE obtained from the linear model. This suggested that the contribution of second order terms in the respiratory model equation was significant, but not major.

This experiment has been reproduced here as a double proof. Also, the plethysmographic and spirometric respiratory signals acquired during this new experiment will permit later in section 11.3.5 the direct comparison of the results of regressive and optimal filter based calibration methods.

.

11.2.1. Experimental protocol

The thoracic and abdominal plethysmographic signals of a subject were simultaneously recorded, along with its spirometric signal, for a period of 30 seconds. The subject was required to remain quiet for 5 minutes prior to the beginning of the acquisition experiment in order for his respiration rate to stabilise. At the same time, corrections were made to the figure-of-8 coil tension and to signal gain and offset to optimise the use of the ADC input span. During the experiment, the subjects were seated on a chair and gently required to breathe normally.

The plethysmographic and spirometric signals were collected by the acquisition computer, and the acquired signals were later processed offline. Not all subjects had exactly the same gain and offset values on their plethysmographic and spirometric signals because it was difficult to reproduce the same settings across all acquisitions. However, the settings were considered reasonably stable.

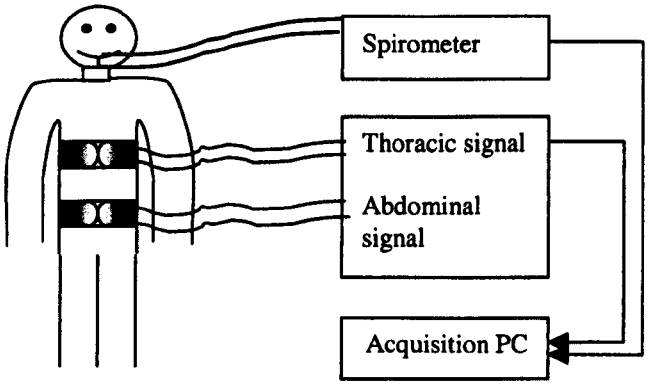

Figure 1. Subject undergoing simultaneous respiratory monitoring with the FORP and a spirometer. The spirometer is used to provide reference signals against which the FORP signals can be compared to.

Four subjects took part of the experiments. There details are summarised in table 1 below.

Subject	Age	Gender	Weight (kg)	Height (m)	Position
1 R.K.	23	Male	95	1.88	Seated
2 L.W.	28	Male	75	1.76	Seated
3 L.C.	28	Male	83	1.75	Seated
4 J.P.R	24	Male	85	1.86	Seated

Table 1. List and vital statistics of subjects involved in the calibration experiment

To calibrate the plethysmographic and respiratory signals, the regressive method was applied to the linear respiratory model (therefore a multiple linear regression was operated) and to the quadratic respiratory model (therefore a multiple quadratic regression was performed).
Specific Matlab scripts were written for this task. The codes for the linear and quadratic regressive method can be found respectively in appendix 7 and 8. The equation for each model is reproduced below. The linear respiratory model is written as follows.

$$P = k_{TH} \cdot \Delta P_{TH} + k_{AB} \cdot \Delta P_{AB} + P_0 \qquad \text{(Eq. 1)}$$

where P is the combined plethysmographic signal, that is to say the virtual respiratory signal, k_{TH} and k_{AB} are the thoracic and abdominal calibration constants, ΔP_{TH} and ΔP_{AB} are the thoracic and abdominal plethysmographic signals, with their mean removed, and finally P_0 is the offset.

The quadratic respiratory model is written as follows.

$$P = l_{TH} \cdot \Delta P_{TH}^2 + k_{TH} \cdot \Delta P_{TH} + l_{AB} \cdot \Delta P_{AB}^2 + k_{AB} \cdot \Delta P_{AB} + P_0 \qquad \text{(Eq. 2)}$$

where P is the virtual respiratory signal, k_{TH} and k_{AB} are the first order thoracic and abdominal calibration constants, l_{TH} and l_{AB} are the second order thoracic and abdominal calibration constants, ΔP_{TH} and ΔP_{AB} are the thoracic and abdominal plethysmographic signals, with their mean removed, and finally P_0 is the offset.

11.2.2. Results and analysis

The results of these calibration operations are reported in table 2 below.

	Linear model				Quadratic model					
Subject	k_{TH}	k_{AB}	p_0	MSE	k_{TH}	l_{TH}	k_{AB}	l_{AB}	p_0	MSE
1	2.143	0.777	3298.888	0.026	2.323	-0.004	0.812	-0.001	3362.792	0.024
2	2.897	0.035	1242.188	0.008	3.815	-0.003	-0.279	0.000	1255.910	0.006
3	5.706	0.808	2276.012	0.012	5.700	0.017	0.821	0.000	2261.359	0.012
4	10.207	0.189	1589.126	0.010	9.942	-0.008	0.823	-0.003	1682.113	0.009
Average	5.238	0.452	N/A	0.014	5.445	0.000	0.544	-0.001	N/A	0.013

Table 2. Summary table of the regression based calibrations on four subjects. In order to make the MSE of different subjects comparable, the spirometric signal of each subject was normalised to fit the range [0; 1]. The offset and gain constants necessary for this transformation were then re-applied to the virtual respiratory signal of this subject. This operation was repeated to all subjects and the MSE was calculated on the difference of the normalised virtual and real respiratory signals.

To help understanding the consequences of these calibrations, the thoracic, abdominal, virtual and real respiratory signals of each subject are now presented. They permit to visually assess the quality of the calibrations.

The real respiratory signal designates the signal measured by the spirometer, whereas the virtual respiratory signal represents the plethysmographic estimation of the spirometric signal.

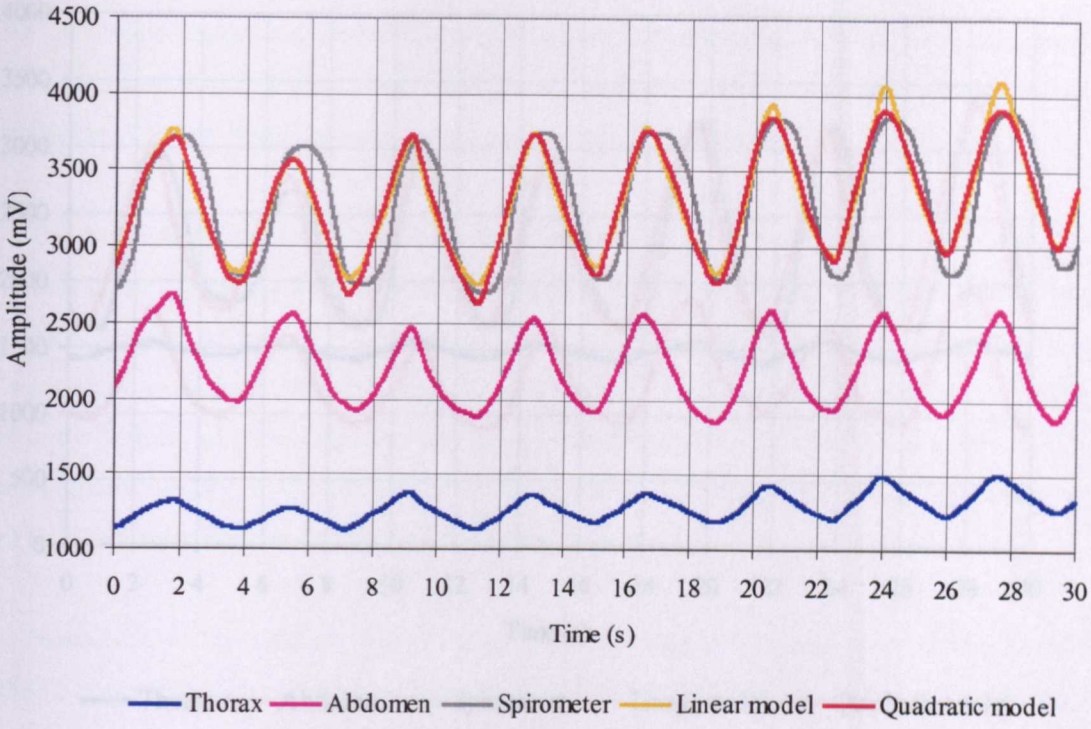

Figure 2. Thoracic, abdominal, virtual and real respiratory signals of subject 1

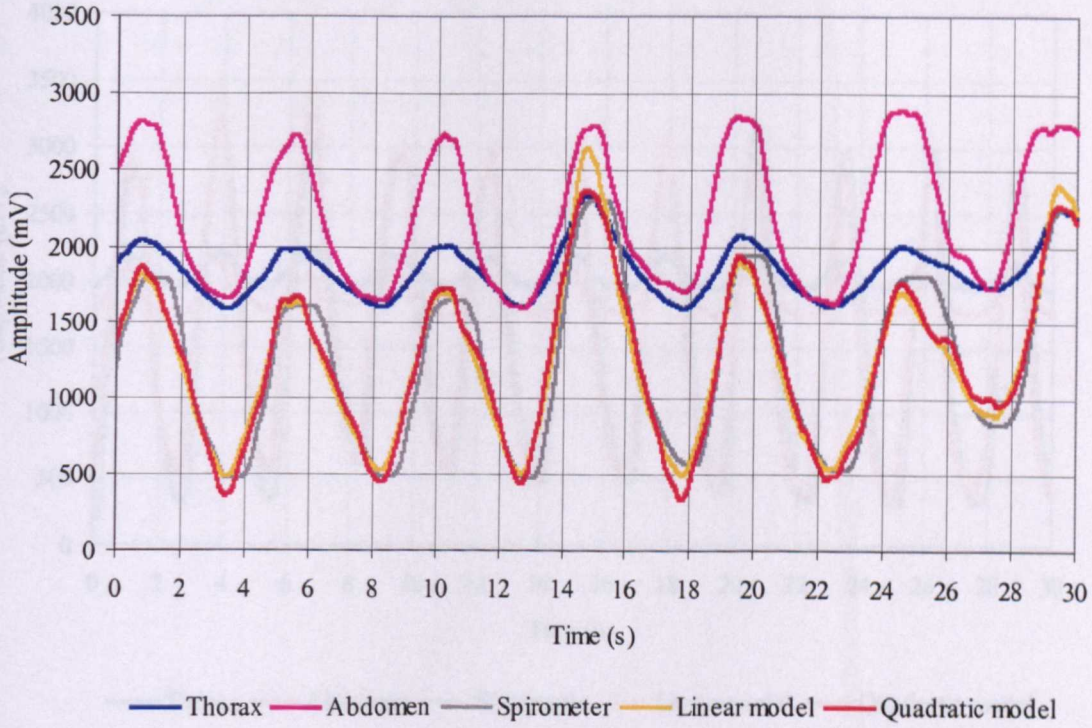

Figure 3. Thoracic, abdominal, virtual and real respiratory signals of subject 2

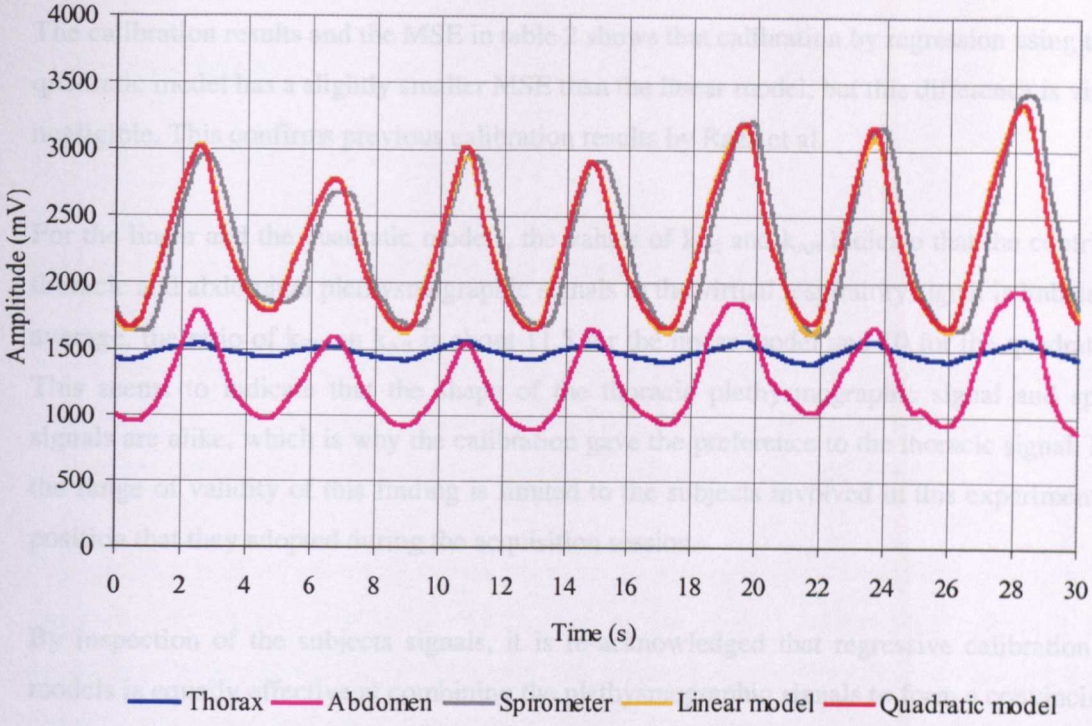

Figure 4. Thoracic, abdominal, virtual and real respiratory signals of subject 3

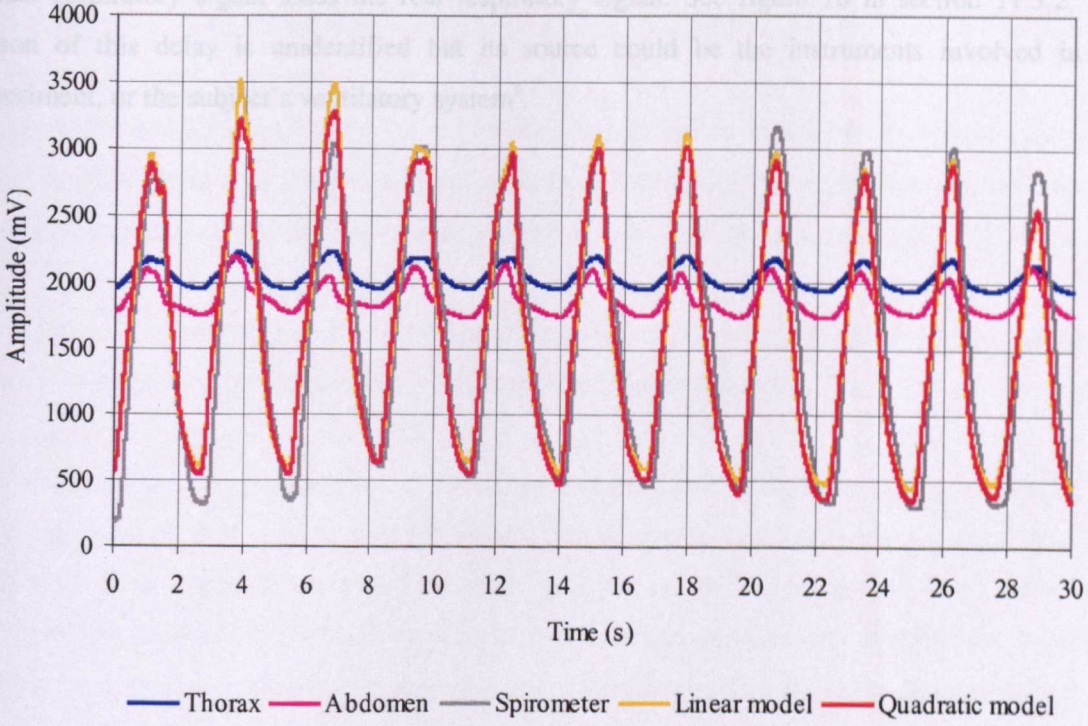

Figure 5. Thoracic, abdominal, virtual and real respiratory signals of subject 4

The calibration results and the MSE in table 2 shows that calibration by regression using the quadratic model has a slightly smaller MSE than the linear model, but this difference is virtually negligible. This confirms previous calibration results by Raza et al.

For the linear and the quadratic models, the values of k_{TH} and k_{AB} indicate that the contribution of thoracic and abdominal plethysmographic signals to the virtual respiratory signal is imbalanced: On average, the ratio of k_{TH} on k_{AB} is about 11.5 for the linear model and 10 for the quadratic model. This seems to indicate that the shape of the thoracic plethysmographic signal and spirometric signals are alike, which is why the calibration gave the preference to the thoracic signal. However, the range of validity of this finding is limited to the subjects involved in this experiment, and the position that they adopted during the acquisition session.

By inspection of the subjects signals, it is re-acknowledged that regressive calibration for both models is equally effective at combining the plethysmographic signals to form a convincing virtual respiratory signal.

An even closer inspection reveals an omnipresent and almost constant time delay between the real and virtual respiratory signals. This delay has a value of approximately 100 to 300 ms and the virtual respiratory signal leads the real respiratory signal. See figure 10 in section 11.3.2. The reason of this delay is unidentified but its source could be the instruments involved in the experiment, or the subject's ventilatory system[1].

[1] The spirometer's arm is balanced with a metal weight to minimise inhalation and exhalation efforts. This mechanical arrangement is equivalent to a damping effect so it is prone to generate delays between the airflow and the volume measurement. The ventilatory system is truly dynamic, given all the elastic elements it contains. Therefore it is also prone to generate delays.
A quantitative separation between the instrumental and physiological effects has not been attempted.
The FORP does not suffer such a large delay: This is because it measures the displacement that causes airflow, as opposed to the spirometer which measures the resulting airflow.

11.2.3. Discussion

11.2.3.1. Is the respiratory model a failure?

While the regressive calibration procedure succeeds at minimising the error between the virtual and the real respiratory signals, such a procedure can also generate surprising calibration values.

It was hoped that, during the calibration process, each signal would be weighted according to the contribution of its respective compartment to the total ventilation volume. In other words, the calibration constants would be positive and their ratio should be in the range of, say, 50 to 200 %.

The generation of negative calibration constants, whose ratio can sometimes be far away from an ideal 100 %, disturbs our belief that the respiratory model should always produce signals in accordance with common sense. In reality, the calibration procedure does not always balance the thoracic and abdominal plethysmographic signals but sometimes uses one signal to modify the other.

These results reveal a hidden assumption that was made when the respiratory models were initially proposed. The models are built from a simplified behaviour of the ventilation mechanics and so the model equations contain mathematical terms (one might say rules), which are meant to produce the behaviour previously observed.
The problem lies with the fact that the calibration process has no awareness of our intentions, neither does it understand the physical meaning of the model's parameters. It simply attempts to solve a set of equations, even if the resulting constants are out of range or negative.

The result of the calibration procedure therefore may no longer reflect the idea that the model contains rules to produce a simplified behaviour of the ventilation system.

This is not necessarily because the included rules were invalid, but because the model did not explicitly contain any expectations of what the calibration constant values should be. The physicality of the model is not questioned here (yet), but rather the interaction of this model with the regressive calibration method. If we accept that the model provides a sufficient degree of physical complexity to reflect the real mechanics of respiration, then the calibration method is not complex enough for the calibration constants to converge to a set of values that would not offend our intuition. For example, calibration constants should be restricted to positive numbers and their ratio should lie within the range, say, [1/3; 3]. However, we might also question the physicality of the model.

11.2.3.2. What is wrong with modelling?

Whatever calibration technique is being used, modelling is based on the fundamental assumption that the thoracic and abdominal plethysmographic signals intrinsically contain the information required to feed the respiratory model. The existence of multiple respiratory models presented in section 7.2, 7.3, 7.4 and 7.5 is based on this assumption but its exact degree veracity cannot be quantified.

Because there is no way to measure this intrinsic veracity, the models were ultimately judged according to their output qualities, i.e. a small error signal and calibration constants that do not offend our expectation.

The calibration results presented earlier in this chapter have shown that the models developed so far have had a limited ability to properly model the ventilation mechanics because the calibration constants could not always fit their expected ranges of values.

The relative failure of this attempt indicates that the models adequacy with the reality of ventilation mechanics has not reached a sufficient level yet, and consequently that more modelling work is required.

11.2.3.3. Is simulating a possible alternative to modelling?

The previous results also clearly indicate that these models can be used as efficient simulations of the spirometric signals, even if doing so was not our initial intention. In the sense of simulating, the results suddenly appear satisfactory.

The whole idea of the respiratory model is now being questioned. Is it so important to remain bound to the approximate mechanical functioning of the ventilatory system? If the results of the calibration are satisfactory, does it matter if the values of the calibration constants are somewhat bizarre?

Being able to produce a good simulation of the real respiratory signal with the FORP obviously would have enormous advantages in the long term non-invasive monitoring of subjects ventilatory activity, in conditions such as sleep, coma, sedation, etc.

So if we now decide to simulate rather than model the real respiratory signal, we suffer no obligation of realism and are free to use any available method to produce such a simulation.

In other words, it is interesting to identify what kind of arbitrary transformation could be applied, if possible at a low computational cost, on the plethysmographic signals to produce a good and consistent simulation.

The underlying assumption of this approach to respiratory monitoring is that the exact relation between the plethysmographic signals and the spirometric signal is undetermined but nevertheless existent and, consequently, both plethysmographic and spirometric signals have somehow related behaviours. Unlike the model approach, we have no interest however in determining a systematic transformation of plethysmographic signals into spirometric signals. Instead, the simulation is recalibrated for each subject, each body position and each style of breathing. The validity of the virtual respiratory signal is then entirely limited to a particular set of conditions, but completely optimised within these conditions.

11.2.3.4. How to generate a simulation ?

An obvious starting point for generating a simulation equation is the regressive calibration of the linear respiratory model.

Among the list of possible corrections to the plethysmographic signals, the existence of a time delay between the plethysmographic signals and the spirometric signal was previously pointed out. By cancelling out this delay, the MSE could be reduced. Additionally, a smoothing filter could be applied to the plethysmographic signal in order to help in mimicking the appearance of the spirometric signal.

A combined solution is to use two separate FIR filters on each plethysmographic signal. The interest in using FIR filters is that, not only can they apply a gain correction to the plethysmographic signals in the same manner as the linear model, but also their phase response can be tailor made, hence providing the required stable time delays. The principle of the modified simulation equation is represented as:

$$P = k_{TH}(jw) \cdot \Delta P_{TH} + k_{AB}(jw) \cdot \Delta P_{AB} + P_0 \qquad \text{(Eq. 3)}$$

where $k_{TH}(jw)$ is the transmittance of the thoracic filter, $k_{AB}(jw)$ is the transmittance of the abdominal filter and jw is the Fourier complex frequency variable. All other terms remain unchanged.

11.3. Wiener optimal filter based calibration procedure

Having defined the structure of a simulation, it is now required to populate the FIR vector of each filter with adequate coefficients. A simple approach to perform this calibration would be to define an algorithm for minimising the mean square difference between the virtual and real respiratory signals. The set of coefficients resulting from such a calibration would define an optimal filter, in the sense of the LMS error.

This calibration procedure is contained within the Wiener-Hopf equation.

The Wiener-Hopf equation, extended to the determination of two simultaneous Wiener optimal filters, is now explained. This extension is termed the Wiener filters linear combiner.

11.3.1. The Wiener filters linear combiner

The Wiener linear combiner is now presented. To start with, the simulation equation, in discreet sample form, in now written as:

$$p_i = \sum_{n=0}^{N-1} a_n \cdot x_{i-n} + \sum_{n=0}^{N-1} b_n \cdot y_{i-n} \qquad \text{(Eq. 4)}$$

where p_i is the virtual respiratory signal, a designates the coefficients of the thoracic filter, x is the thoracic plethysmographic signal, b designates the coefficients of the abdominal filter, and y is the abdominal plethysmographic signal. Both filters have the same size N.

In matrix form, this same equation is written as:

$$p_i = A^T X_i + B^T Y_i \qquad \text{(Eq. 5)}$$

where A is the thoracic impulse response vector containing the N coefficients a_n of the thoracic filter, B is the abdominal impulse response vector containing the N coefficients b_n of the abdominal filter, X_i is a column vector containing the last N values of the thoracic signal x, and Y_i is a column vector containing the last N values of the abdominal signal y. The superscript ^{T} designates a matrix transposition.

The difference between the virtual and real respiratory signals is termed the error signal e_i. Such a signal is defined as:

$$e_i = s_i - p_i \qquad \text{(Eq. 6)}$$

where s_i is the real respiratory signal, as measured with a spirometer. Next, the mean square error J is defined as:

$$J = \frac{1}{M}\sum_{i=0}^{M-1} e_i^2 = E[e_i^2] \qquad \text{(Eq. 7)}$$

where M is the duration in samples of the error signal. The operator E designates the average (or expectation in this case). The development of J leads to the following equation:

$$J = E[s_i^2] - 2E[s_i X_i^T]A - 2E[s_i Y_i^T]B + A^T E[X_i X_i^T]A + B^T E[Y_i Y_i^T]B + 2A^T E[X_i Y_i^T]B \qquad \text{(Eq. 8)}$$

It is acknowledged that J is a second order function of A and B. Therefore, J will admit a unique minimum J_{opt} for optimal values of the impulse responses A and B, namely A_{opt} and B_{opt}. In order to find A_{opt} and B_{opt}, we must solve the following system:

$$\begin{cases} \dfrac{\partial J}{\partial A} = 0 \\ \dfrac{\partial J}{\partial B} = 0 \end{cases} \qquad \text{(Eq. 9)}$$

This system of two simultaneous equations is expanded and re-arranged to give:

$$\begin{cases} E[s_i X_i^T] = A^T E[X_i X_i^T] + B^T E[Y_i X_i^T] \\ E[s_i Y_i^T] = B^T E[Y_i Y_i^T] + A^T E[X_i Y_i^T] \end{cases} \qquad \text{(Eq. 10)}$$

This system is also equal to:

$$\begin{cases} P_{sx}^T = A^T R_{xx} + B^T R_{yx} \\ P_{sy}^T = A^T R_{xy} + B^T R_{yy} \end{cases} \qquad \text{(Eq. 11)}$$

where P_{sx} and P_{sy} designate column vectors of correlation coefficients, while R_{xx}, R_{xy}, R_{yx} and R_{yy} designate a square matrix of correlation coefficients. In transposed form, this system is also equal to:

$$\begin{cases} P_{sx} = R_{xx}A + R_{xy}B \\ P_{sy} = R_{yx}A + R_{yy}B \end{cases} \qquad \text{(Eq. 12)}$$

Note that the transposed form of R_{xy} is R_{yx}, and vice versa. Expressed in "matrix of matrix" (or tensor) form, this system of two simultaneous matrix equations can be further reduced to:

$$\begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \cdot \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} P_{sx} \\ P_{sy} \end{bmatrix} \quad \text{(Eq. 13)}$$

The above tensor equation is then expressed as:

$$R \cdot C = P \quad \text{(Eq. 14)}$$

where R is the square correlation coefficients tensor, P is the column correlation coefficients tensor, and C is the tensor containing the impulse responses of both filters.
To find the optimal impulse responses A_{opt} and B_{opt}, the tensor C must be determined. By tensor inversion, we obtain the following solution:

$$C = R^{-1} \cdot P \quad \text{(Eq. 15)}$$

The above equation is similar in form to the Wiener-Hopf equation, with the difference that it was implemented using tensors in order to produce two sets of filter coefficients.

The necessary code for solving this extended Wiener-Hopf equation was implemented in Matlab and is reproduced in Appendix 9. The computation of correlation coefficients for populating the R and C tensors, and the R tensor inversion were greatly assisted by built-in operators and functions.

The whole duration of plethysmographic and spirometric signals (30 seconds) was included in the calculation of A_{opt} and B_{opt}.
To ensure a large safety margin the production of the required delays, the duration of the impulse response was set to 1 second, and therefore N = 2000.

The respiratory waveforms are relatively simple generally, and present a sinusoidal aspect when the subject is at rest. In such conditions, the respiratory frequency should be around 0.5 Hz and the frequency range is between 0 and 2 Hz. In other words, the filter's frequency properties above 2 Hz should not have effects on the input signals.

11.3.2. Results and analysis

Once the optimal thoracic and abdominal filters were determined for each subject, the characteristics of the filters were determined, the virtual respiratory signal was generated, and the MSE was calculated. These results are summarised below in table 3.

	Optimal filter					
Subject	k_{TH}	τ_{TH} (ms)	k_{AB}	τ_{AB} (ms)	p_0	MSE
1	1.826	132	1.209	-478	3299	0.006
2	2.293	-22	0.433	-696	1242	0.001
3	8.159	-285	0.728	-188	2276	0.001
4	8.971	-174	1.957	162	1589	0.003
Average	5.313	-87	1.082	-300	N/A	0.003

Table 3. Summary table of the results of the optimal filter based simulation of the real respiratory signal. The constants k_{TH} and k_{AB} correspond to the low frequency magnitude of each filter. The constants τ_{TH} and τ_{AB} correspond to the low frequency phase delay of each filter. These constants are representative of the filter behaviour in the bandwidth of the respiratory signal (because this bandwidth is limited to 2Hz). The MSE of each subject was calculated with the same normalisation procedure as for the regressive calibration.

The real respiratory signal is the signal obtained from the spirometer. The virtual respiratory signal is the plethysmographic estimation of the spirometer signal.

The real and virtual respiratory waveforms (obtained by regressive calibration and optimal filtering) of each subject are now displayed for comparative purpose.

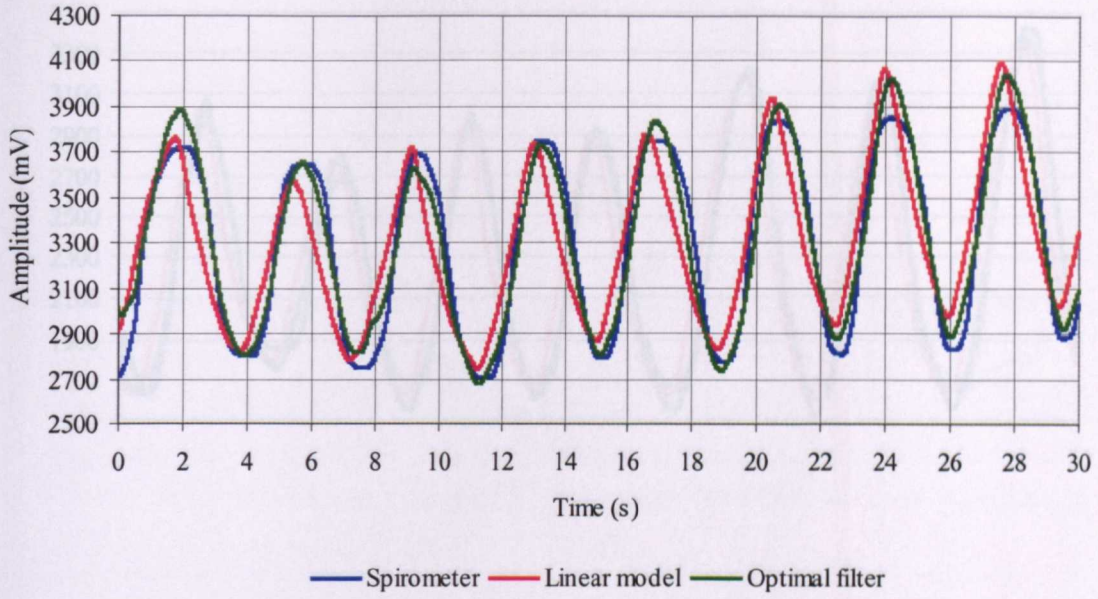

Figure 6. Spirometer signal compared to virtual respiratory signals obtained by regressive calibration for the linear respiratory model and optimal filtering for subject 1.

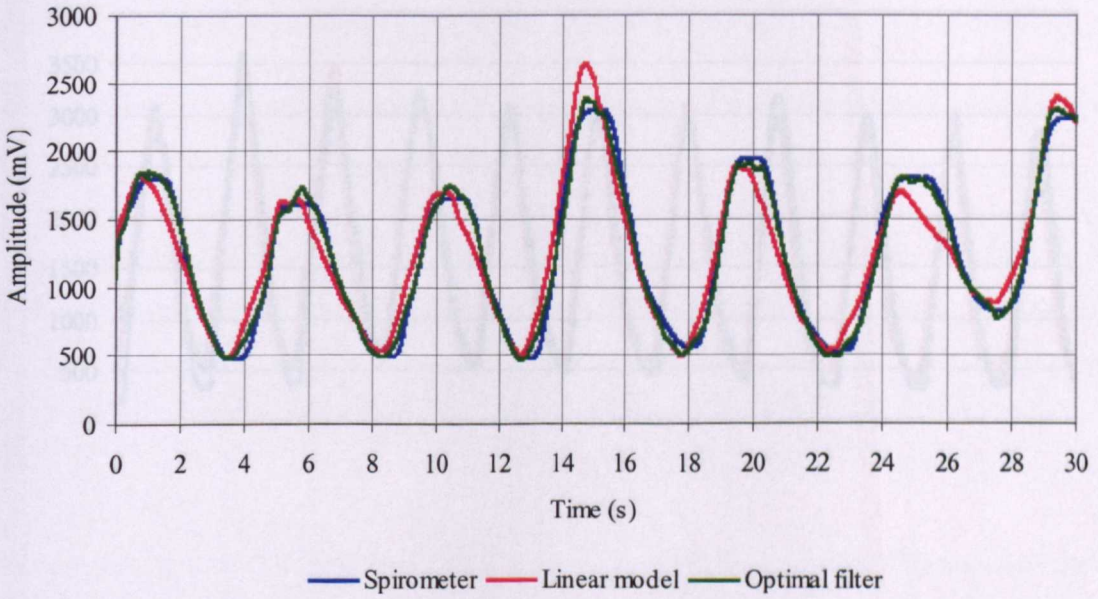

Figure 7. Spirometer signal compared to virtual respiratory signals obtained by regressive calibration for the linear respiratory model and optimal filtering for subject 2.

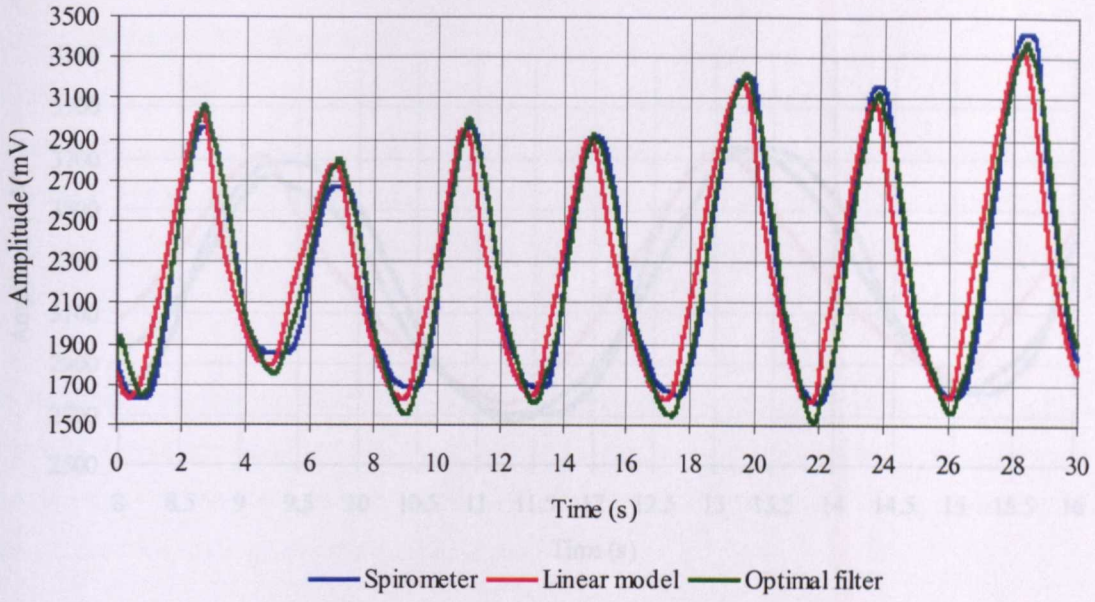

Figure 8. Spirometer signal compared to virtual respiratory signals obtained by regressive calibration of the linear respiratory model and optimal filtering for subject 3.

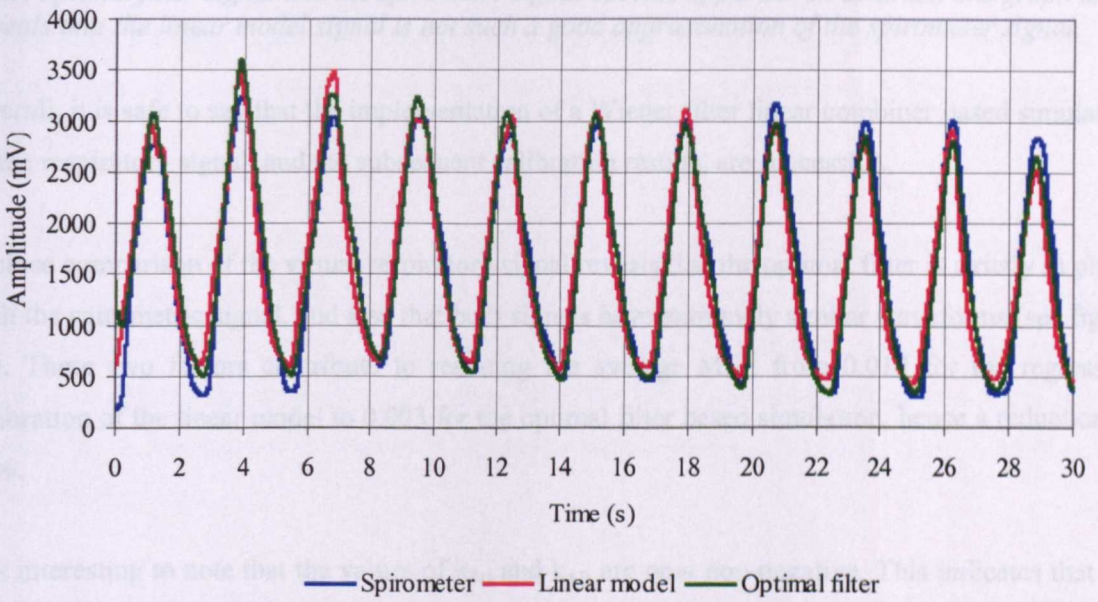

Figure 9. Spirometer signal compared to virtual respiratory signals obtained by regressive calibration for the linear respiratory model and optimal filtering for subject 4.

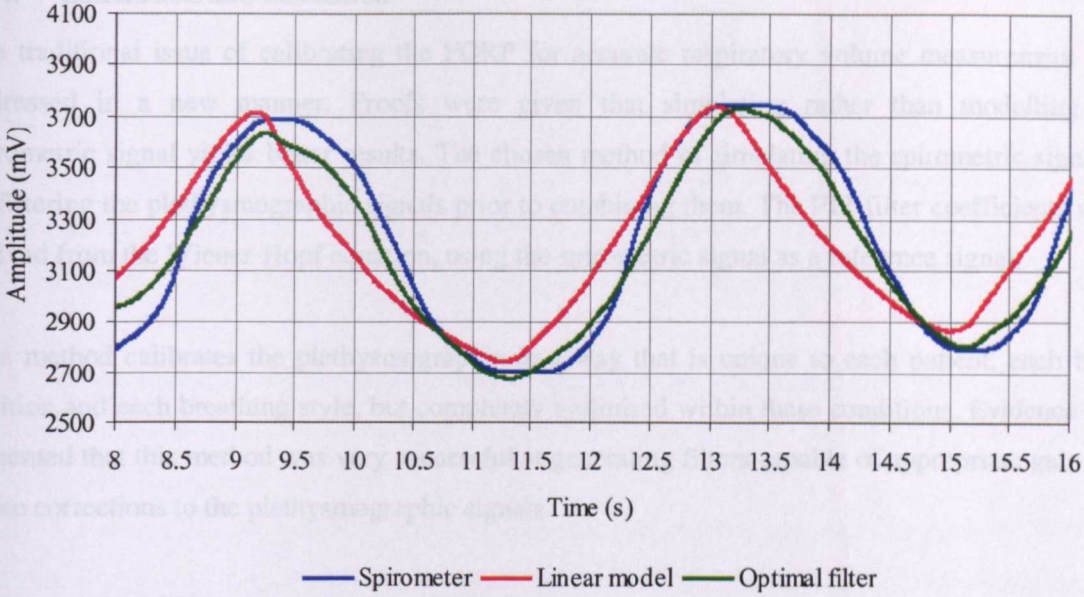

Figure 10: Enlargement of figure 5 between 8 and 16 seconds. The similarities of shape and phase of the optimal filter signal and the spirometer signal become apparent. In contrast, this graph also reveals that the linear model signal is not such a good approximation of the spirometer signal.

Overall, it is safe to say that the implementation of a Wiener filter linear combiner based simulation of the respiratory signal, and the subsequent calibration results, are successful.

A close comparison of the virtual respiratory signal reveals that the optimal filter is strictly in phase with the spirometric signal, and also that both signals have extremely similar waveforms (see figure 10). These two factors contribute to reducing the average MSE from 0.014 for the regressive calibration of the linear model to 0.003 for the optimal filter based simulation, hence a reduction of 79%.

It is interesting to note that the values of k_{TH} and k_{AB} are now non-negative. This indicates that the filtered plethysmographic signals are used in a constructive, rather destructive, manner to form the virtual respiratory signal.

It is also interesting to note that the average of k_{TH} remains larger than the average of k_{AB}, which somehow confirms to the idea that, within the conditions of this experiment, the spirometric signal has a stronger correlation with the thoracic signal than the abdominal signal. However, the ratio k_{TH}/k_{AB} is now around 5, which about 50% smaller than with a regressive calibration.

11.4. Conclusion and discussion

The traditional issue of calibrating the FORP for accurate respiratory volume measurement was addressed in a new manner. Proofs were given that simulating rather than modelling the spirometric signal yields better results. The chosen method of simulating the spirometric signal is by filtering the plethysmographic signals prior to combining them. The FIR filter coefficients were derived from the Wiener-Hopf equation, using the spirometric signal as a reference signal.

This method calibrates the plethysmographic in a way that is unique to each patient, each body position and each breathing style, but completely optimised within these conditions. Evidence was presented that this method was very successful at generating filters capable of appropriate gain and phase corrections to the plethysmographic signals.

The normalised MSE was reduced by 79%, compared to the classic regressive calibration of the linear respiratory model.

Interestingly, the initial regressive calibration of the linear respiratory model can now be regarded as a particular case of the Wiener filter based calibration, when the size of each filter is reduced to a single coefficient.

The Wiener filter linear combiner employed here used two filters. However, such a structure could in theory possess as many FIR filters as required to accept input signals. It is therefore envisaged that another two inputs, corresponding to the squared thoracic and abdominal signals, could be added, thus effectively introducing non-linearity in the filters combiner. While one might fear to increase the dependence of the results on experimental conditions, it is likely that the MSE could be even further decreased.

The validity of the simulation was not tested past the boundary of the experiment so it remains to be determined how often the simulation needs re-calibration, depending on the frequency and the amplitude of the subject's non-ventilatory related movements and change of position. Each time a major change is detected, new set of coefficients will be required to carry on delivering accurate simulations. Assuming that the subject remains quiet, a first assumption is that the longer the duration of the calibration, the longer the simulation is likely to remain valid.

11.5. Further work

11.5.1. Possible real time implementation

The work presented in this chapter is only a proof of principle and the system speed was not a critical parameter. However, a further proof of principle might be interested in realising a real time implementation of the system. Such implementation would require specific developments.

After the determination of the coefficients of the thoracic and abdominal filters, the virtual respiratory signal is obtained by convolution of the plethysmographic signals with their respective Finite Impulse Responses.
Given that each FIR contains N = 2000 coefficients, (1s impulse response duration * 2000 Hz sampling frequency), the convolution requires about 4000 multiplications and 4000 additions to produce a single new value of the virtual respiratory signal. In total, a real time implementation of the linear combiner would require a processor capable of handling about 2000*(4000 + 4000) = 16 millions of operations per second.

A reduction of the sampling frequency to, say, 100 Hz would subsequently reduce the number of operations per seconds to 40000, which any modern (or less modern) processor could handle.
In particular, there is scope for an embedded implementation of the system, using a PIC microcontroller, or maybe a DSP chip for more computational power.
Because the plethysmographic signals have such a low frequency domain, such a reduction of the sampling frequency would not significantly degrade the quality of the acquired signals.

Finally, one might consider a reduction in the number of bits required to represent a discrete sample of the signal to speed up calculation.

In this condition, one can say that, once the filters calibration has been performed, the virtual respiratory signal could be generated in real time with a standard desktop computer. This remains to be tested.

11.5.2. Possible immediate improvements

The use of Wiener optimal filter for the simulation of spirometric signals is successful, but at a high computational cost during the calibration period. This translates into long calculation time. In the present system, 30 seconds of signals history was accumulated prior to calibration. Then, about 30 seconds of calculation time[2] was required to determine the optimal filters coefficients. In total, the

[2] These calculations were executed on an AMD K6-II processor at frequency 333 MHz. A more recent processor would offer a smaller calculation time.

system has a dead time of about 1 minute before being ready to start generating the virtual respiratory signal. In clinical conditions, this dead time might be unacceptable.

Decisive improvements to the present system would therefore consist of 1) optimising the Matlab script for calculation speed, 2) possibly reducing the amount of signal history required prior to calibration, and 3) decreasing the computational burden associated with the inversion of the correlation tensor R, and its product to the correlation tensor P.

The second issue is answered simply: it is conceivable that the amount of signal history required prior to calibration could be reduced if the subject's breathing pattern remained reasonably constant. For example, a quiet patient may only require 20 seconds of signal accumulation if the breathing pattern shows a reasonably good ergodicity. The third issue could also be simply answered by decreasing the sampling frequency (as mentioned in section 11.5.1) and increasing the computational power. However, a more radical solution would be to simplify (and therefore accelerate) the process of determining suitable filters coefficients. This possibility is discussed in section 11.5.3.

11.5.3. Possible alternative to the Wiener filter

The Least Mean Square (LMS) adaptive filter [Widrow et al. (1975)] presents itself as a logical step in the direction of simplifying the calibration procedure by adopting a recursive approach to the problem of determining optimal filters coefficients. This is now explained in more details.

The LMS adaptive filter avoids the heavy matrix manipulations described in equation 15 by estimating (as opposed to determining) at low computational cost the filters coefficients contained in the vector C of equation 15.
Such estimation is based on a recursive process called the steepest descent algorithm that progressively converge to the optimal filters coefficients. The rate of convergence can be controlled, and slow convergence means better stability and accuracy of the optimal coefficients estimation.

Like the Wiener optimal filters, two separate LMS adaptive filters can be arranged to work together in a linear combiner.

Once the coefficients have converged to relatively stable values, they are considered as good estimations of the coefficients generated with an optimal filter, in the same conditions. It is therefore expected that the virtual respiratory signal generated by the LMS adaptive filter would present the same characteristic of minimal MSE than the virtual respiratory signal generated by an optimal filter. When the user is confident that the adaptive filter has converged to a suitable set of

coefficients (this can be measured with the MSE), the virtual respiratory signal can replace the spirometer signal. The calibration period required of the adaptive filter corresponds to the time taken by the coefficients to converge to their expected values. This duration depends on two factors: The rate of convergence parameter and the stationarity of the plethysmographic and spirometric signals. Stationary signals will lead to a faster convergence that non-stationary ones for the reason that a change in the characteristics of any of the three signals forces the adaptive filter to aim for a different set of optimal coefficients. This re-adaptation is not instantaneous.

A preliminary test using an LMS adaptive filters linear combiner for generating a virtual respiratory was not conclusive: The rate of convergence of the adaptive filters was too fast and the impulse response duration was too short. As a result, the filter coefficients were never converging but always re-adapting and, consequently, the system was incapable of producing a stand-alone estimation of the spirometric signal.

However, it is strongly possible that a longer impulse response (at least equal to the longest breathing cycle of the subject) and a smaller convergence rate would permit to reach a stable set of coefficients that would make possible the generation of an accurate virtual respiratory signal at low computational cost. This remains to be tested.

Chapter 12 CARDIAC MONITORING WITH THE FORP DURING MRI AND CT INVESTIGATIONS

12.1. Introduction

The new version of the FORP, which was developed during this research program, provides sufficient degree of details in the signal to observe cardiac activity to the limit of simple visual inspection. It was suggested in chapter 8 that cardiac monitoring with the FORP could potentially be extended to usage in the MR or CT scanners during cardiac and/or cardiovascular imaging.

Both systems need a precise source of synchronisation when imaging the heart or the circulatory system and the position of the heart is currently estimated from a parallel ECG recording. However, the ECG is not completely reliable in this matter because of phantom peaks in the signal due to magnetised blood flow [Roth (1996)]. An alternative or complementary method of synchronisation would benefit the imaging quality of both scanner types. Producing truly accurate synchronisation of the CT or MR scanners at any time of the cardiac cycle would require knowledge of the heart position with an instrument safe and compatible with both scanners, i.e. non-metallic, non-invasive and operating in real time.

It appeared that such non-ECG based synchronisation method by cardiac motion tracking could be provided by the thoracocardiogram (TCG) signal of a plethysmocardiographic instrument such as the FORP or the RIP. The TCG signal contains information on heart volume, and more precisely ventricular volume if the sensor is correctly placed on the chest. Analysis of the TCG signal would permit precise localisation of early as well as late systolic phases, a task currently difficult with the ECG system because the ECG trace is flat during end systole, and contains no information on heart position. Early systole is perceived by the TCG as a maximum ventricular volume and late systole as a nearly flat ventricular filling curve.

In practice, the RIP could neither be used within the MR scanner nor with the CT scanner because its sensor consist of a metallic wire. In the CT scanner, the periodic pattern of the electrical wire implemented inside the sensor belt would interfere with the imager in the region where the sensor is situated. This is because metal opacity to X-rays would prevent the radiations emanated by the CT scanner from penetrating or leaving the body slice encircled by the sensor. This implies in particular that cardiac imaging could not be achieved since the sensor would have to be positioned at the heart's level, just below the xiphoid process. In the case of the MR scanner, the RIP could potentially be used if the sensor could be built of non-magnetic materials. However the sensor would act as an antenna to the magnetic field variations and conduct induced currents to the RIP machine, resulting in very noisy measurements. There is also a risk that induced currents might heat up the sensor to the point of burning the patient.

In contrast, the FORP sensor is safe and compatible with both MR and CT scanner technologies since optical fibres are 1) insensitive to magnetic field variations and 2) relatively transparent to X-rays. In addition, health and safety risks associated with its use are almost non-existent. Perhaps more importantly, a previous version of the FORP had already been used successfully in different MRI suites for respiratory gating during long examinations. [Raza (1998)]
It was therefore proposed to investigate the possibility of using the FORP for heart position monitoring with the TCG signal in view of synchronising the scanner (either CT or MR) during cardiac and/or cardiovascular imaging.

It was also proposed to study whether cardiac gating with the FORP could potentially replace the ECG gating system in the MR scanner. This is for two reasons. Firstly, use of the ECG system presents a health and safety issue: the periodic gradient magnetic fields produced by the MR scanner coils can induce electrical currents in the ECG cables if these cables are crossed or form a loop, thus resulting in cables heating up and potential skin burns[1] [Gosbee and DeRosier (2001)].
Secondly, in MRI conditions the triggering is rendered difficult because the ECG signal obtained in situ is very noisy and often difficult to interpret due to 1) artefact currents in the ECG cables generated by the gradient magnetic fields and 2) phantom peaks in the ECG signal created by natural flow of magnetised blood. Artefact currents are limited by using high impedance ECG cables but ultimately special processing is required to clear up the signal prior to triggering[2].

Additionally, the TCG signal contains information on the subject's position: detecting faint body displacement and discarding corresponding cardiac cycles should contribute to better imaging.

In the same manner as the ECG, synchronisation of the MR and CT scanners to the subject's heart movements with a TCG signal could be obtained with by different methods, either by prospective triggering or by retrospective gating. These methods are explained below.

---

[1] This can be a serious issue if the patient has been sedated and is incapable of alerting the scanner operator. Skin burns due to crossed or looped ECG cables are the second cause of MRI accidents, after the projectile effect of magnetic materials. To limit skin burns, the ECG leads must be kept apart and must be heavily shielded, except of course at the extremities where they are connected to the electrodes.
[2] The precise form of this processing is not made known by MRI equipment companies, probably for commercial reasons. It is thought to be some kind of bandpass filtering.

12.1.1. Prospective triggering

Prospective triggering is a real-time procedure, taking place during acquisition. In practice, an ECG recording of the patient is acquired in the scanner and each time the R wave is detected, the scanner is triggered after a delay specified according to what phase of the cardiac cycle is desired for imaging. The trigger delay is usually calculated as a percentage of the most recent R-R interval duration, the most popular value being 80%. The R wave is traditionally chosen because its short-lived high amplitude peak is simple to detect algorithmically, and also because it corresponds to ventricular depolarisation, which is the event initiating ventricular contraction, hence cardiac movement. Like the ECG, the TCG signal has algorithmically recognisable features: a maximum (corresponding to maximum ventricular volume during early systole) and a minimum (corresponding to minimum ventricular volume during late systole). During prospective triggering, the ECG-derived trigger signal must be generated in real time.

The duration of the whole process of data acquisition with ECG gating varies greatly, from sub-second in Electron Beam CT (EBCT) scans where only one heart cycle is sufficient to construct an image, to a few minutes in older MR scans where hundreds of heartbeats are necessary for imaging. In such slower scanners, the construction of an image is divided into the acquisition of image layers, the layer acquisition rate being one per cardiac cycle.

The method of delayed ECG triggering makes the assumption that, at a constant heart rate, the position of the heart is consistently the same for a constant delay after the R wave. Consequently, at a varying heart rate, a particular position of the cardiac cyle can be reached by scaling the delay according to the heart period. However, this is only an approximation. [Bazett (1920)] In fact, a previous plethysmographic study shows that, even at a constant heart rate, there is a natural variability in the interval between the R wave and a particular phase of the cardiac cycle, as detected by the RIP [Jordan et al. (1984)]. In other words, prospective ECG triggering is more accurate for early systolic cardiac phases (close to the R wave) than for late systolic phases because imaging of systolic late phases requires a long delay during which the uncertainty on the exact heart position grows. Consequently, data acquired during late systole with the ECG trigger might contain slightly different phases of the cardiac cycle and this results in blurring while constructing the image, hence degradation of performance.

Evolved ECG triggering methods involve the use of a knowledge base that corrects the trigger delay according to the instantaneous heart rate and a model of the heart dynamics [Sanchez-Ortiz and Burger (1996)]. The resulting time delay is presumably more accurate but still only estimated.

The prospective triggering is best suited for scanners with low temporal resolution but the acquisition of a layer must be shorter than one cardiac cycle.

The condition imposed on the FORP system by the MR and CT scanners during prospective gating would be that the TCG is extracted in real time, and possibly without the use of a reference ECG signal. Such real time, standalone analysis of the TCG has had no previous report.

12.1.2. Retrospective gating

Retrospective gating is a procedure where the ECG signal is not employed during the acquisition to trigger the scanner, but after the acquisition to indicate time regions containing a static heart position. This type of gating also permits the creation of "cine loops" where a complete heart cycle can be imaged and looped to create the illusion of cyclic movement. Obviously, this is only possible if the temporal resolution of the scanner is very high so that multiple image layers can be acquired during a single cardiac cycle, as opposed to prospective triggering which was used for scanners limited to the acquisition of one layer per cardiac cycle. Retrospective gating is possible with MR scanners using fast pulse sequences[3] and multislices imaging[4], or with EBCT scanners. The ECG signal does not need to be processed in real time but offline at a later moment, so retrospective gating permits more effective and precise R wave detection because more time can be dedicated to this task and also because not just the past but the entire signal history is known at the moment of the analysis.

The aim of this chapter is to show examples of PCG signals obtained with the FORP and to study the feasibility of using the TCG signal as an alternative, or additionally, to the ECG signal for scanner gating/triggering. This work essentially concerns the development and analysis of specific digital signal processing techniques.

12.2. Experimental methods

The acquisition experiences have been performed on multiple subjects with no previous record of cardiac or respiratory disorders, all capable of holding their breath without chest convulsion for half a minute but untrained to maintain their chest at a stable position over the apnoea period. The aim of this experience was to record FORP signals for latter signal processing algorithm development and testing. Prior to acquisition, each subject was instructed on the program of the experiment. Each experiment consisted of a variable number of 30 second long acquisition sessions during which the subject was required to breathe or to hold his/her breath while remaining in the supine position.

---

[3] Fast pulse sequence techniques are now well developed. The principle is to rotate the proton magnetic vector by only a few degrees in order to obtain a faster recovery time. This allows the reduction of the repetition time of the pulse sequence, but at the expense of the total acquisition time. See chapter 6 for explanation of the MRI principles.

[4] In multislices imaging, the MR signals from multiple interlaced slices can be acquired at once, thanks to simultaneous gradient fields.

The FORP sensor was connected up at the location of interest to pick up a specific PCG. Most experiments consisted of recording the TCG signal, i.e. the PCG signal obtained when the sensor is positioned at or below the xiphoid process.

A 3-leads ECG signal obtained in lead II configuration (see section 3.3.1) was also recorded for time reference purposes simultaneously with the FORP signals. ECG electrodes were connected to an electrically isolated ECG pre-amplifier (Harvard Apparatus).
The subject was required to remain quiet for 5 minutes prior to the beginning of the acquisition experiment in order for his/her heart and respiration rates to stabilise. At the same time, corrections were made to the figure-of-8 tension and to the FORP and ECG signals gain and offset to optimise the use of the ADC input span.

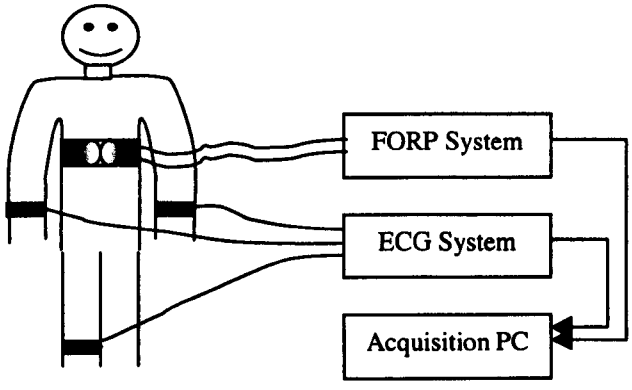

Figure 1: Subject undergoing cardiac monitoring with the FORP.

Male and female subjects could be recorded equally easily since the localisation of the FORP sensor for TCG acquisition (below the xiphoid process) did not require an important level of intimacy, as opposed to the positioning of ventral or dorsal ECG electrodes, prior to MRI or CT examinations.

The TCG signals of three subjects will be studied. The subject description list is given below in table 1. Due to the small number of subjects involved in this study, the results are only preliminary, indicating a direction for further research. It is believed however that further testing on other subjects should not produce significantly different signals, given the strong resemblance of the ones we present here.

Subject	Age	Gender	Weight (kg)	Height (m)	Position
1 L.O.	27	Male	73	1.71	Supine
2 F.B.	25	Male	69	1.74	Supine
3 D.C	30	Male	78	1.65	Supine

Table 1. Subjects description table

12.3. Type of PCG signals obtained from the FORP

The TCG signal is the only signal containing information on the heart volume but not the only cardiac signal that could be measured by the FORP. In fact a variety of sensor locations can be used to generate PCG signals, including abdomen, wrists, arms, neck and head.

Thoracic, abdominal, neck and head signals are now presented in conjunction with a simultaneous ECG signal as a visual proof of results. The signals were obtained from subject 1, in a standing position.

The ECGs and the PCGs were treated offline with a band pass Butterworth filter and a zero phase filtering algorithm (the "filtfilt" function in the Matlab DSP toolbox). This ensured that the filtered ECG and PCG signals remained exactly in phase with each other and with the unfiltered signal, therefore permitting easy comparison.

The high pass side had a 1 Hz cut-off and order 2 slope, whereas the low pass side had a 20 Hz cut-off, also with an order 2 slope. The zero phase filtering algorithm operates by filtering the same sequence twice, back and forth. This not only exactly compensates the original filter's phase difference but also doubles the order of the original filter. See appendix 2 for the proof. The Butterworth filter was chosen for its flat spectral response at the cut-off frequencies.

The signals to be represented had to be down-sampled from 2 kHz (the sampling frequency of the acquisition system) to 200 Hz. This led to a decrease by a factor of 10 in the size of the spreadsheet file used for creating the graphs, without impairing the visual quality of the signals.

After down-sampling, both PCGs and ECGs were scaled and positioned to fit on the same graph. The head PCG was given the same treatments with the difference that the low pass filter cut off frequency was 3 Hz instead of 30 Hz. This is because the original signal noise level was extremely high.

For measurement of head and neck PCGs, the sensor's neoprene support belt had to be removed (because it was too long) and the optical fibre was attached directly to the subject.

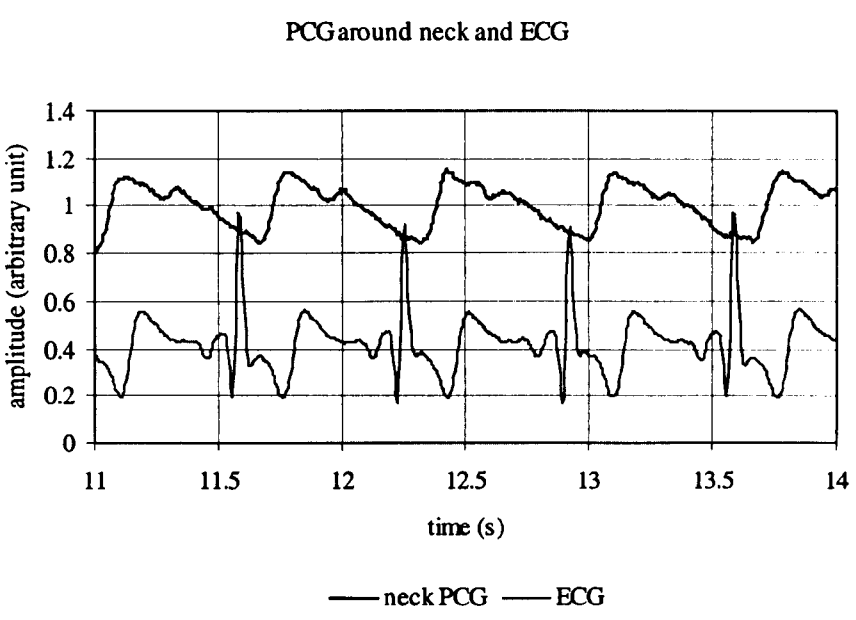

Figure 2: Typical PCG signal with the sensor placed around the neck during respiration and simultaneous ECG recording. The ECG and the PCG signals are bandpass filtered in between 1 and 30 Hz. The subject was seated and remained quiet while holding his breath during the investigation. When the signal is acquired during ventilation, the amplitude of the respiratory component remains faint. Therefore the PCG signal acquired around the neck has a potential for robust cardiac rate measurement. It could easily be envisaged to use such signal in MR or CT scanners during cardiac imaging to verify the TCG results. More straightforwardly, it could be used to synchronise the scanner during imaging of the neck. The origin of the signal is due presumably the displacement of a large volume of blood close to the surface of the skin, very probably in the carotid artery. A partial proof is that the PCG signal shown above closely resembles the carotid pressure waveform obtained by applanation tonometry [Kelly et al. (1989)].

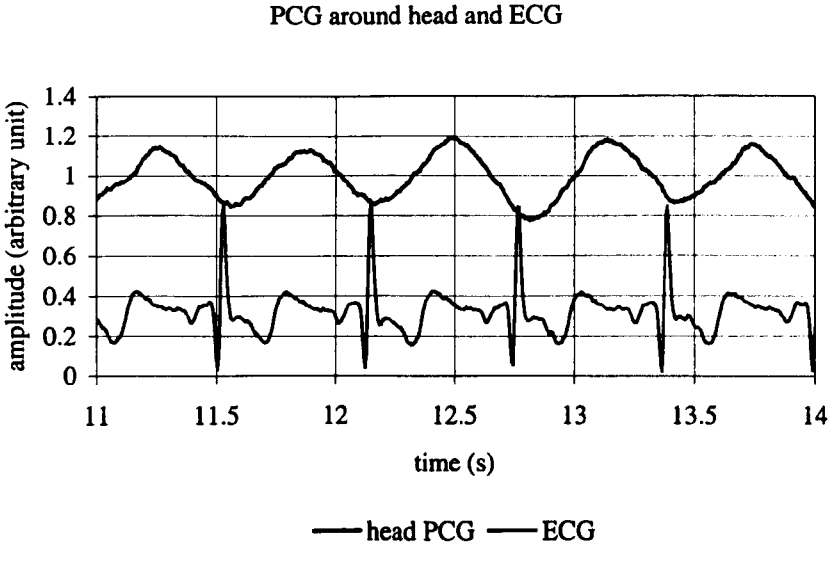

Figure 3: Typical PCG signal around head during respiration and simultaneous ECG signal. The PCG signal is band pass filtered between 1 and 3 Hz and the ECG signal is band pass filtered between 1 and 30 Hz. Note the sinusoidal character of the PCG signal due to the very small cut-off frequency of the denoising filter. The subject was seated during the investigation. The signal presumably originates from volume variation of skin superficial blood vessels, presumably in the temporal area where the heartbeat can be felt with a fingertip.

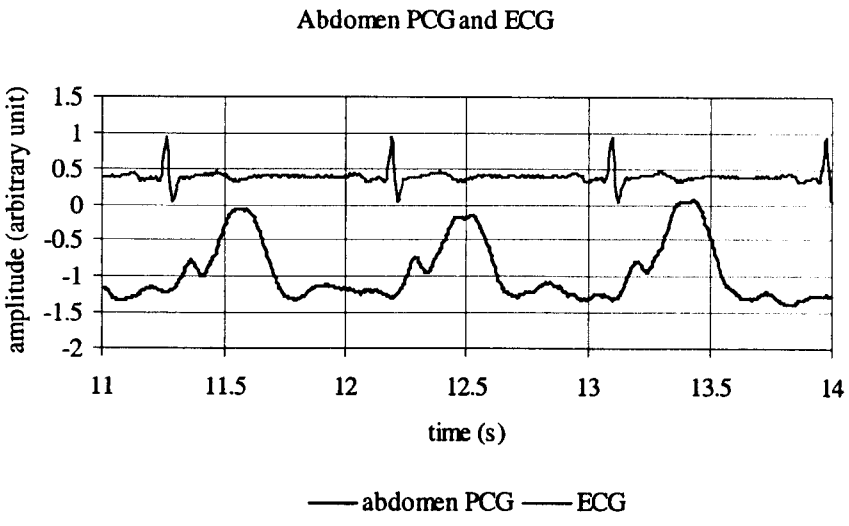

Figure 4: Typical PCG around abdomen during apnoea and simultaneous ECG signal during breath-hold. The PCG and the ECG signals are band pass filtered between 1 and 30 Hz. The subject was seated and held his breath during the investigation. The sensor was located around the abdomen and was aligned with the navel. The signal is presumably under the influence of both cardiac movements and local abdominal blood flow: The large peak is probably due to local blood volume variations and consequently, the smaller peak ahead of the large one is probably due to ventricular contractions.

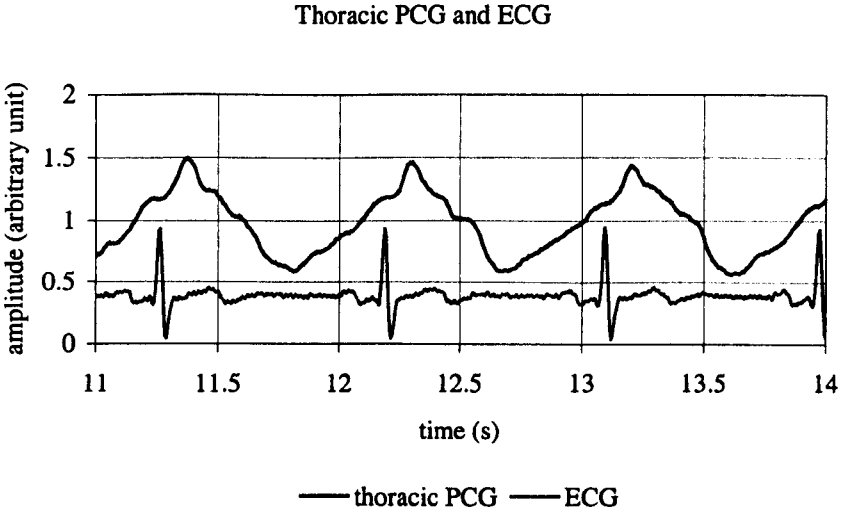

Figure 5: Typical Thoracic PCG (preferably called TCG) and simultaneous ECG signal during breath-hold. The TCG and the ECG are bandpass filtered between 1 and 30 Hz. The sensor was placed 3 cm below the xiphoid process. The subject was seated during the investigation. The TCG signal originates from the displacement of the heart due to its contractions. Despite damping due to lower mechanical impedance of the lungs, this variation in heart volume transmits through pulmonary tissues to produce a chest volume variation.

12.4. Study and algorithmic analysis of the TCG signal

The TCG signal, as observed during apnoea, is now studied in detail. The study aims at determining if the TCG is a possible alternative to the ECG trigger in cardiac imaging with MR or CT scanners.

Five criteria emerged from searching for the optimal trigger point in the TCG signal. They are listed below in order of importance:

- The TCG signal must contain an anchor, i.e. an easily recognizable, shape invariant event of high repeatability. The anchor position should serve as a time reference within each cardiac cycle[5].
- The TCG signal must contain a "stable" region of minimal first derivative and maximal repeatability across all cardiac cycles, during which the image's layer can be acquired.
- The duration of this region should be at least equal to the scanner aperture time, in order to limit motion blurring.
- The time delay between the anchor and the beginning of the stable region should be statistically significant so that the location of the stable region could be accurately estimated from the location of the anchor. This would permit limiting the algorithmic analysis to the detection of the anchor only.
- The stable region should also be as close as possible from the anchor to limit progressive de-synchronisation of the two events due to heart rate variations.

Even if only the first criterion were satisfied, the TCG signal would already present itself as an alternative to the ECG trigger. In this case, an anchor could be used instead of the R wave in any triggering method originally designed for use with the ECG. For example, one can think of, say, a "80% max to max" method, in the same fashion as the classic "80% R to R" method mentioned above.

However, if all the criteria listed above can be satisfied, then localising the stable regions in the TCG signal could well be regarded as a potentially better source of synchronisation for cardiac imaging than methods employing an ECG signal.

Such a trigger algorithm based on the localisation of stable regions in the TCG would therefore need knowledge of 1) how to detect the anchor and 2) how to detect the stable region with respect to the anchor.

---

[5] An anchor manifests itself as the firing of a particular event indicating the completion of one cardiac cycle. In comparison with the ECG signal, the R wave is an anchor thanks to its regularity of shape and its relative independence from the rest of the population of ECG events.

The remainder of this study of the TCG signal is aimed at attempting to prove that the signal actually contains anchors and flat regions that can be detected in an algorithmic manner, either offline or in real time, with or without knowledge of a parallel ECG recording.

In the first part of the study (see section 12.4.1), a suitable anchor is determined simply by carefully examining the TCG signal. The signal is then processed offline in Matlab 1) to detect the anchor with a specifically designed algorithm, 2) to produce an analysis of the anchor detection results, and finally 3) to statistically determine the most stable region in the signal. At that point, sufficient knowledge of the TCG signal has been gathered to perform an "all-in-one" algorithmic analysis where the raw signal can be effectively processed at once to determine an optimal trigger point for each cardiac cycle within the signal. Such offline processing would be involved with scanners operating in retrospective gating mode.

The second part of the TCG study (see section 12.4.2) is concerned with determining whether the same "all-in-one" analysis could be performed in real time with the same level of accuracy as its offline counterpart. This second part aims at producing an algorithm that would be used with scanners operating in prospective triggering mode. Again, all algorithms have been developed with Matlab and written in a way specific to the offline simulation of a real-time process. They have not been tested in "true" real time conditions.

Each part of the study (offline and real-time) is sectioned into ECG assisted and ECG independent analysis.

12.4.1. Offline study of TCG signals

It was decided to perform signal processing in the time domain because the frequency domain approach is inappropriate: cardiac and respiratory signals have slow cycles and for that reason most of their harmonic content is spread in the range of 0 to 10 Hz. Such low frequencies cannot be properly resolved with a spectral representation[6].

The acquisition system sampling frequency is set to 2 kHz, which is unusually high for a biological signals acquisition system but provides a very good signal precision after denoising thanks to the large quantity of samples produced per second. Consequently, all operations require more intensive calculation but this is of no importance as far as offline work is concerned.

---

[6] The frequency resolution (the smallest frequency variation that can be resolved with spectral analysis) being the inverse of the signal duration, a modest 1 BPM spectral separation would require a 60 second long acquisition, which is obviously too long for the patient in apnoea.

During apnoea, respiratory movements do not interfere with the TCG signal[7] but the latest still has to be separated from other components such as a large time varying DC signal due to spurious chest displacement[8] and noise of analogue and digital origins. The first step in localising anchors is to design a band pass filter to de-noise and de-trend the raw signal.

12.4.1.1. Filtering

A band-pass Butterworth digital filter is employed to clean up the signal and remove its trend. The filter is constituted of an order 6 low pass stage of cut off frequency F_{low} equal to 20 Hz, and an order 2 high pass stage of F_{high} equal to 1 Hz.

The value of F_{low} was chosen as a compromise between noise attenuation and loss of signal resolution. By careful examination of signals filtered in different manners, 20 Hz seemed to be a valid boundary. The order 6 ensures a strong filtering effect thanks to an attenuation of 6* 20 = 120 dB per decade, or 6 * 6 = 36 dB per octave.
The value of F_{high} was chosen to produce a valid boundary between the signal and its trend. The high pass attenuation is 40 dB per decade frequency, or 12 dB per double frequency.

The Butterworth criterion (maximally flat response at the cut off frequency) was chosen to produce a clean separation between the pass band and the lateral stop-bands but this choice was not critical and another classic filter, such as the Chebyshev, would probably produce similar results.

The coefficients of the digital Butterworth filter for the low pass and high pass stages were generated separately using the "butter" function in Matlab. In essence, the butter function generates the coefficients of an analog transmittance function corresponding to the requirements given by the user, and the transmittance function is converted to the digital domain using the bilinear transformation. See section 5.14.2.4.
One inconvenience of the Butterworth criterion is that it hinders the filter with a particularly non-linear phase response across the pass-band, therefore occasioning phase distortion in the filtered signal.

To eliminate phase distortion, a phase cancellation filter algorithm has been used (the "filtfilt" function in the Matlab DSP toolbox). The filtfilt function filters the same signal back and forth,

---

[7] The respiratory component is difficult to separate from the cardiogenic TCG component because of their large amplitude difference in a co-existent frequency range. It is estimated that chest circumference variation due to cardiac activity at xiphoid level is sub-millimetric because the sensor displacement during apnoea is usually invisible to the naked eye. In contrast, quiet ventilation provokes chest circumference variation of a few centimetres. The maximum contribution of the cardiogenic component to the respiratory component is about 5%.
[8] It is difficult, even for a concentrated and willing subject, to maintain a stable chest position over half a minute in apnoea. Consequently the signal tends to drift.

resulting in an exactly zero phase delay, and doubling the original filter order. (The proof is given in appendix 2) Effectively, the resulting filter has order 2*(2 + 6) = 16.

The same bandpass filter algorithm is applied to the ECG.

12.4.1.2. Normalising

Different acquisition sessions generate signals with different amplitude and offset. It is important to establish some standardisation in gain and offset between all signals so that further processing algorithms do not depend on the signal scale and position. In particular, the anchor detection algorithm (see below section 12.4.1.5) requires that its input signals conform to a specific amplitude range because part of this algorithm is based on constant level thresholding.

Additionally, the TCG signal may contains localised scale variations due to spurious chest movements that could not be filtered (because their spectral region match with the passband) and which should be minimised prior to anchor detection.

There are numerous ways to normalise a signal and a common statistical manner is to subtract its average and divide the difference by its standard deviation.

However, by replacing the average and standard deviation by a moving average and a moving standard deviation, we can not only correct the overall signal's amplitude range but also compensate for localised spurious dispersions and hopefully minimise them.

In practice, there is no need for subtracting the moving average from the filtered signal because the "near zero" frequency components have already been removed by the high pass side of the filter. As a consequence, the standard deviation is equivalent to the square root of the moving average of the squared filtered signal. The moving average is implemented using a non-causal, phase cancellation filter (see appendix 2).
This normalisation is a non-linear process where the signal is modulated by the inverse of its own standard deviation. This produces dynamic compression (less gain) or expansion (more gain) of the input signal. This is somehow related to a high pass effect in the sense that this process will not allow a signal to remain "flat", in the sense of a small standard deviation.

When deciding what is an appropriate behaviour for the normalisation process, the duration of the moving standard deviation is a critical parameter. If it is too long, no local normalisation effect is observed and the normalised signal could still contain spurious deviations. If it is too short, the local normalisation will perceive the signal as if through a magnifying glass, and this will result in underestimating the standard deviation. In this case the normalisation process will attempt to distort

the signal to increase its standard deviation, therefore eradicating any distinctive shapes to replace them by oscillations. This raises the question of determining where is the boundary between "correcting" and "transforming" the original signal.

Experimentally, the moving standard deviation has been computed over 1 second. This duration corresponds to the normal heart rate of 60 BPM after a few minutes of relaxation. This prevents the normalisation from trying to temper with the TCG waveform.
Once normalised, all TCG signals are expected to show ergodicity[9] and to systematically fit in an amplitude range of ± 3 arbitrary units.

12.4.1.3. Spectral analysis

The spectral analysis of the TCG signals from the three subjects is now represented. The interest in these representations is in understanding the harmonic structure the TCG signal.
The Power Spectral Density (PSD) were all calculated with the Welch modified periodogram (see section 5.11.4). Each TCG signal contained 65536 samples, and the PSD was obtained by averaging 7 FFTs of consecutive and overlapping regions of the signal. Each region was first weighted with a Hamming window, had a size of 16384 samples, and the overlap constituted 8192 samples, i.e. 50% of each region.

In the Welch method, the advantage of a large overlap over a large number of regions is to introduce a statistical dependence between the FT of all the regions, therefore decreasing the variance of the PSD.

Of course, slicing the original signal into regions limit the spectral resolution of the PSD but this is really acceptable given the large number of points in each region (16384).

Each PSD was generated from the Graphical User Interface (GUI) of the signal processing toolbox "sptool" in Matlab, then transferred to Excel and truncated above 50 Hz to concentrate on the frequency region of interest, and subsequently limit the size of the Excel file.

[9] A signal has ergodicity if the average of all of its periods is equivalent to the average of one period.

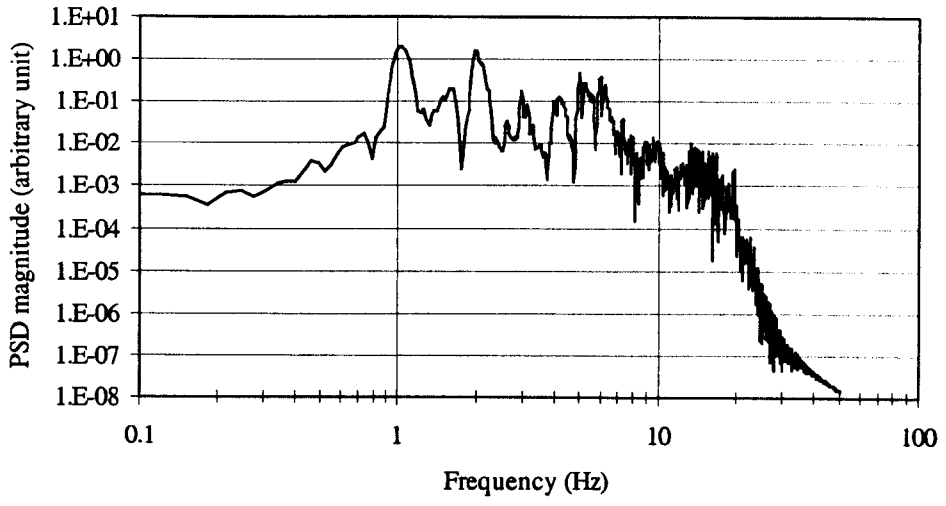

Figure 6: Welch modified periodogram of subject 1.

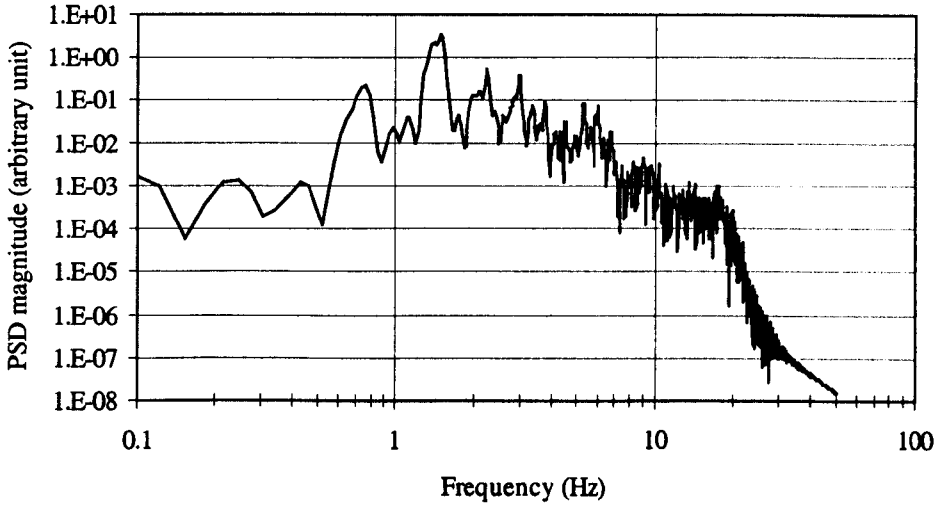

Figure 7: Welch modified periodogram of subject 2.

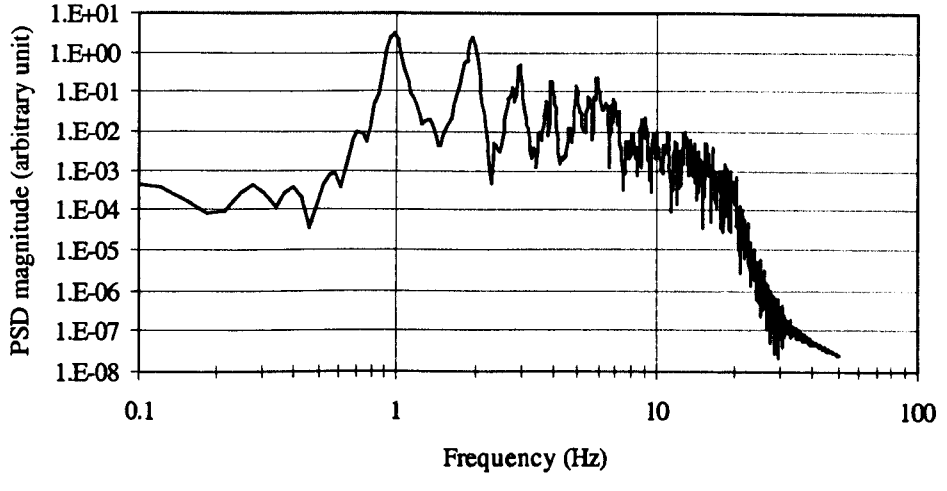

Figure 8: Welch modified periodogram of subject 3.

The three spectral representations above show a strong density of both even and odd harmonics. The harmonic population is located between 1 and 20 Hz, as a consequence of previous filtering. The large plateau below 1 Hz indicates a large number of sub-second, aperiodic components in the TCG signal, which could not be eliminated by the filter.

12.4.1.4. Analysis of waveform

After filtering and normalisation, the TCG signal is ready for display. We are particularly interested in assessing 1) the consistency of each signal across its cardiac cycles and 2) the resemblance of all signals.

A simple evaluation technique of TCG consistency is to inspect all the signals over their whole duration. However, it would be difficult to reproduce all the signals on paper with a scale large enough to resolve important details.

To avoid the problem, the TCG signals were sliced into time windows of duration 2 s, centred on the R wave of a parallel ECG recording. Each window was then in phase with a particular R wave and consequently all windows were in phase. Next, the TCG slices were superimposed and then displayed. Because of the superimposition, this representation technique provides a significant economy of space. This technique also provided a way to visually assess the repeatability of a signal.

The average signal of the TCG signal across all slices was calculated. The average signal shows a template TCG waveform. It is expressed below:

$$MEAN(t) = \frac{1}{N_{slices}} \sum_{n=1}^{N_{slices}} slice(t,n) \qquad \text{(Eq. 1)}$$

where MEAN(t) is the average signal of all slices at instant t, N_{slices} is the total number of slices that were cut down from the original TCG signal, and slice(t,n) is an array containing the TCG signal in slices.

In the same fashion, the standard deviation signal was calculated. The standard deviation signal measures the dispersion from the average signal. At any time of the cardiac cycle, the lower is the standard deviation signal, the higher is the consistency of a particular phase. It is expressed below as:

$$SD(t) = \sqrt{\frac{1}{N_{slices} - 1} \sum_{n=1}^{N_{slices}} [slice(t,n) - MEAN(t)]^2} \qquad \text{(Eq. 2)}$$

where SD(t) is the standard deviation of all slices at instant t.

One limitation of this visualisation technique however is in the signification of the average signal and the standard deviation signal. Away from the time origin, the slices progressively lose their phase coherence (due to heart rate variability) and consequently the average and standard deviation signals become less statistically significant.

Finding appropriate anchors is simple: They must be extrema with a high repeatability and amplitude. Such points are easily located from the superimposed slices of TCG signal.

Finding appropriate stable regions where the scanner could acquire a layer with minimal motion blur is more challenging. A stable region is perceived as both flat and consistent. In other words, such region must have a minimal first derivative of the average waveform and a small standard deviation. A stability estimation method is proposed to attempt localising stable regions according to the two criteria mentioned above. First, a stability estimation function, termed Stability Estimator (SE), is generated. To produce a combined assessment of the flatness and the consistency, the SE was built of the average (over the scanner aperture time) of the absolute value of the first derivative of the average signal, weighted by the standard deviation signal. The SE function is expressed below:

$$SE(t) = \frac{1}{T_{ap}} \cdot \sum_{\tau=0}^{T_{ap}} SD(t+\tau) \cdot \left| 1 + \frac{dMEAN(t+\tau)}{dt} \right| \qquad \text{(Eq. 3)}$$

where T_{ap} is the aperture time of the scanner, SD(t) is the standard deviation of the TCG slices at instant t, MEAN(t) is the average of the TCG slices at instant t. We took the absolute value of the first derivative because we are interested in measuring the quantity of variation, rather than the direction of variation. The term 1 inside the vertical bars is included to ensure that there is not an excessive dependence of the standard deviation on the first derivative of the average, the latter having intrinsically greater variability than the standard deviation.

Secondly, we look for the minimum of the SE function. When a minimum is found in t_{min}, we can say that the TCG signal provides maximal stability (in the sense of the SE function) in the region starting from t_{min} and finishing at $t_{min} + T_{ap}$. On the graphs below, the SE function has been calculated with an aperture time T_{ap} of 100 ms, which is the worst case and should suit slow scanners.

If the SE function exhibits multiple valleys of similar depth instead of a single minimum, three parameters should contribute to making a decision on which valley to choose as the beginning of the best stable region. The ideal valley should be close to the R wave, it should have the longest duration possible, and it must be as deep as possible.

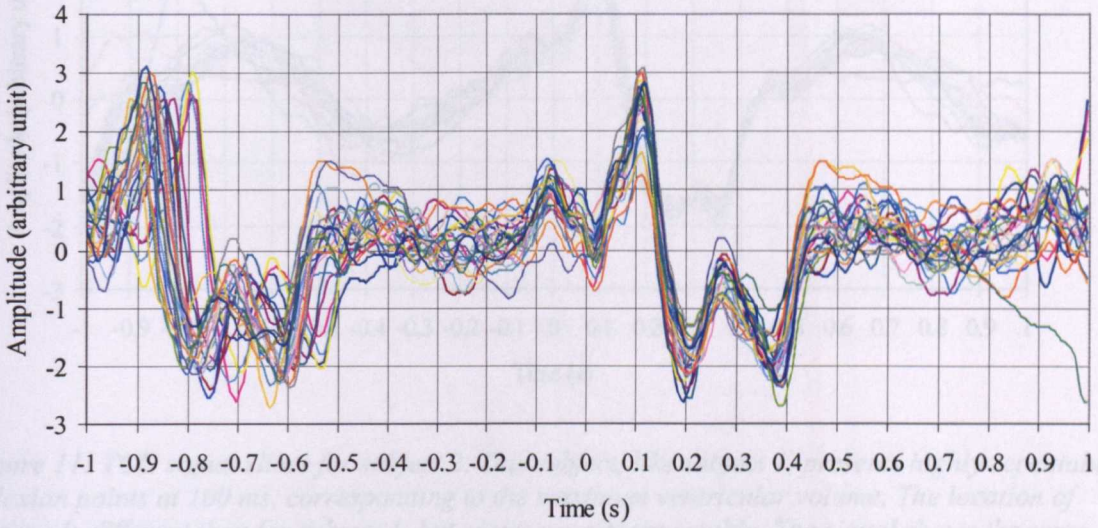

Figure 9: TCG signal slices for subject 1. The TCG signal was cut down in a time interval of ± 1s around the R wave of the parallel ECG recording. The resulting slices have been superimposed for representation. We can see immediately that the inflexion points situated between 100 and 400 ms after the R wave are candidate anchors for this subject. After 600 ms, the signal is stabilising.

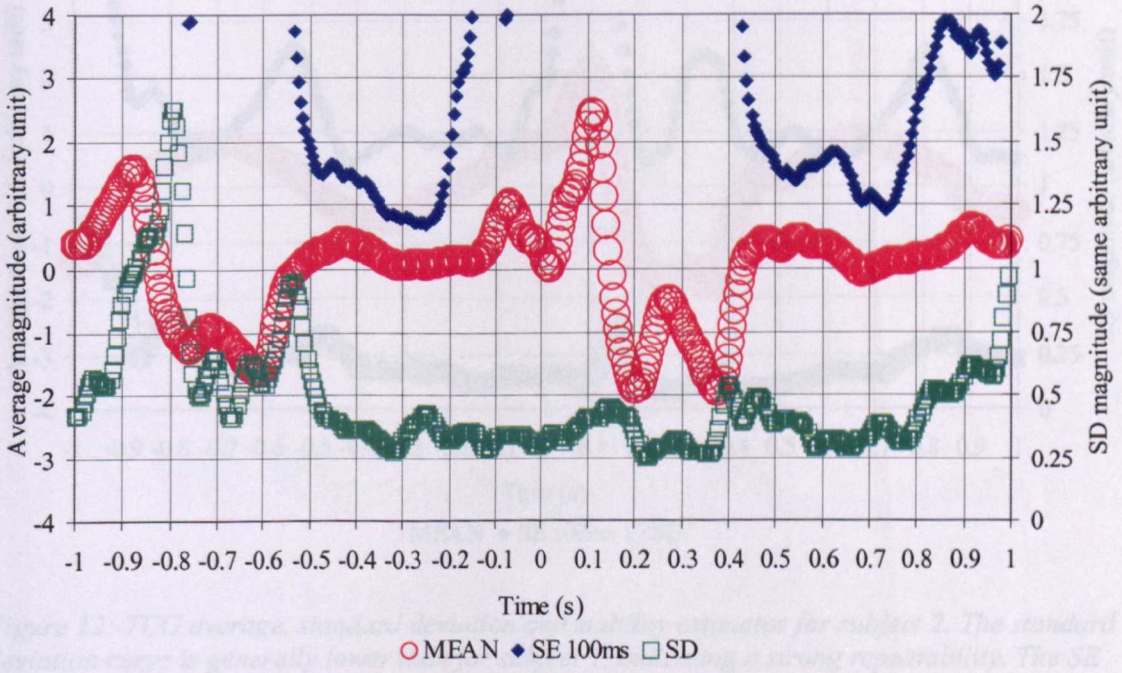

Figure 10: TCG average (mean), standard deviation (SD) and stability estimator (SE) for subject 1.The SE function indicates two distinctive minima, at 740 ms and –250 ms. This early minimum is better because it is deep, large and close to the R wave. However, it is not accessible in real time.

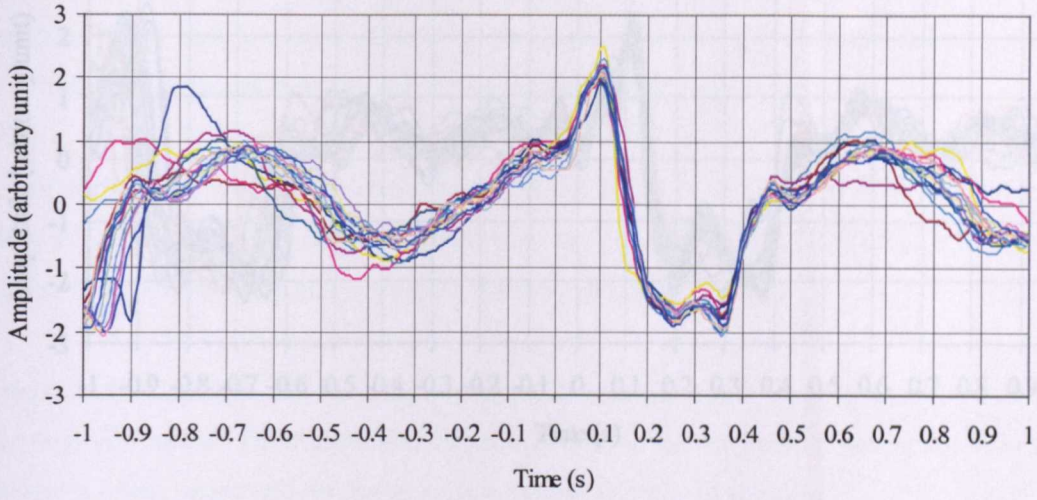

Figure 11: TCG signal slices for subject 2. This subject, like subject 1, presents highly repeatable inflexion points at 100 ms, corresponding to the maximum ventricular volume. The location of minima is different than for subject 1, but seems equally repeatable. The signal shows the same amount of dispersion on both sides of the time axis, indicating a small heart rate variation

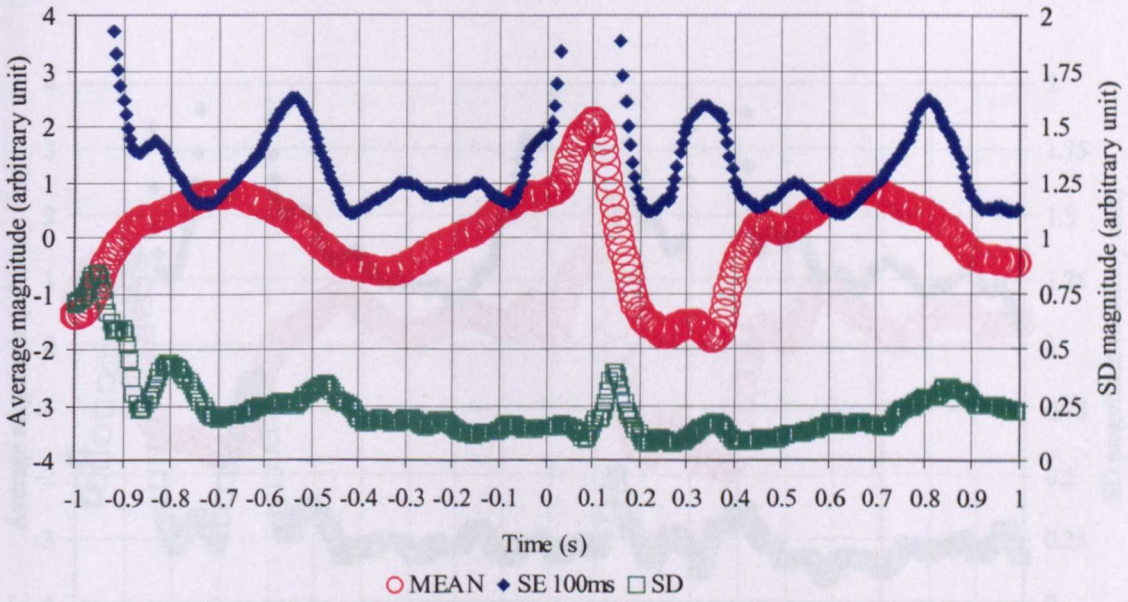

Figure 12: TCG average, standard deviation and stability estimator for subject 2. The standard deviation curve is generally lower than for subject 1, indicating a strong repeatability. The SE function contains multiple minima of similar amplitude, therefore signifying the co-existence of different possible trigger points. On the positive side of the time axis, 450 ms seems to be the best location, because the SE minimum is large, in association with a low standard deviation. Minima of the SE function situated before null time are equivalent in amplitude to those after null time, confirming a small heart rate variability for this subject.

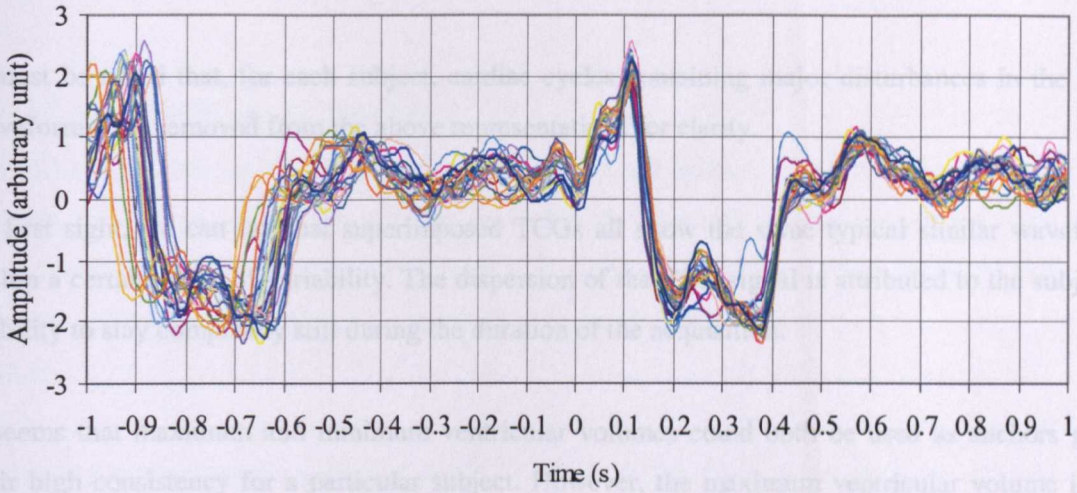

Figure 13: TCG signal slices of subject 3. The maximum ventricular volume, located at 100 ms, seems as consistent as for the previous two subjects. Yet again, the two consecutive minima exhibit a subject dependent behaviour. The signal seems more stable before the R wave than after, which corresponds to a rapid loss of phase coherence, due to heart rate variability. For proof, observe the maxima at –900 ms.

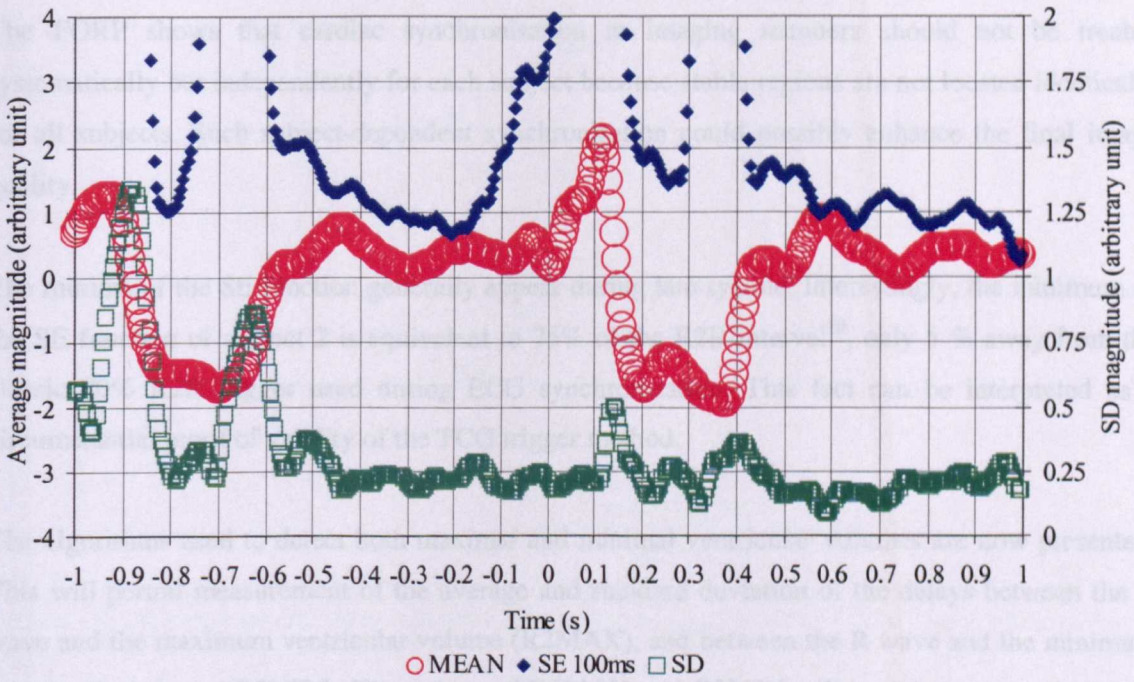

Figure 14: TCG average, standard deviation and stability estimator of subject 3. The SE function shows two equivalent minima at 650 and 800 ms. It is unclear which of them could be the best trigger point. We can see another minimum at –200 ms.

It must be noted that, for each subject, cardiac cycles containing major disturbances in the TCG waveform were removed from the above representations for clarity.

At first sight, we can say that superimposed TCGs all show the same typical similar waveform, within a certain range of variability. The dispersion of the TCG signal is attributed to the subject's inability to stay completely still during the duration of the acquisition.

It seems that maximum and minimum ventricular volumes could both be used as anchors given their high consistency for a particular subject. However, the maximum ventricular volume is the only event whose shape and timing remains consistent across all subjects.

Unlike the ECG, the TCG allows us to observe the natural variations of cardiac motion in different subjects. Cardiac function is, of course, invariable in nature but the heart of each subject seems to present a specific variation of the pattern, like a signature. This should not be a surprise if we consider that the heart is mostly made of muscular layers whose development is intrinsically related to the general physical activity and physiognomy of the body, hence differing from one subject to another.

The FORP shows that cardiac synchronisation in imaging scanners should not be treated systematically but independently for each subject because stable regions are not located identically for all subjects. Such subject-dependent synchronisation could possibly enhance the final image quality.

The minima of the SE function generally appear during late systole. Interestingly, the minimum of the SE function of subject 2 is equivalent to 75% of the R2R interval[10], only 5 % away from the classic 80% R2R trigger used during ECG synchronisation. This fact can be interpreted as a circumstantial proof of validity of the TCG trigger method.

The algorithms used to detect both maximal and minimal ventricular volumes are now presented. This will permit measurement of the average and standard deviation of the delays between the R wave and the maximum ventricular volume (R2MAX), and between the R wave and the minimum ventricular volume (R2MIN). The values of R2MAX and R2MIN will permit an appreciation of the consistency of both events and finally lets us conclude which of them should be used as an anchor.

[10] The heart period of subject 2 is stable and close to 1000 ms. Its optimal trigger point is 750 ms away from the R wave, which corresponds to 75% of the R to R interval.

12.4.1.5. Algorithmic anchor detection

The TCG signal having been denoised, detrended and normalised, it will now be shown how the position of extrema ventricular volumes are detected. This task is decomposed into two major steps: firstly extrema must be found and marked (or "flagged") and secondly, erroneous flags must be discarded so that only flags corresponding to maximum and minimum ventricular volumes remain.

The same process must be applied to the ECG signal to flag the R wave, with the difference that only the first step is necessary in this case. This is because the R wave, given its very distinctive amplitude, is unmistakable. (This process retrospectively explains how the TCG signal was sliced out according to the position of each R wave: a slice was produced for each ECG flag encountered.)

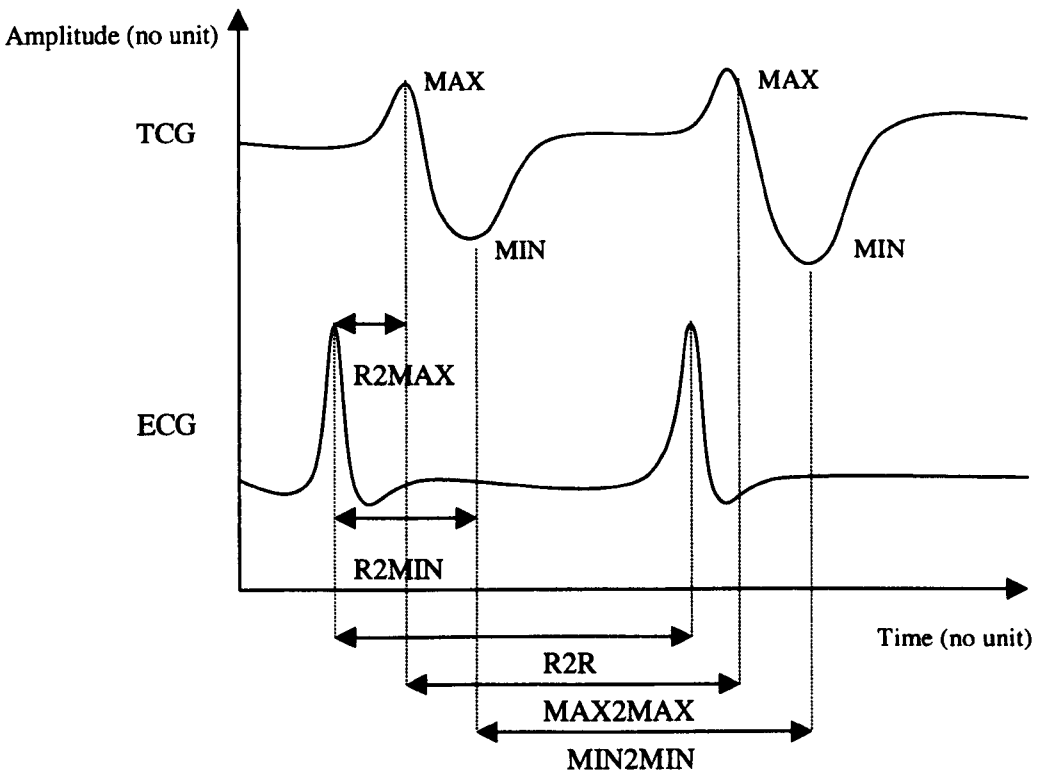

Figure 15: Different specific time intervals relative to the ECG and the TCG signals. Qualitative graph.

12.4.1.5.1. Detecting extrema of TCG signals and generating flags

The position of each extremum lying above (detection of maximal ventricular volume) or below (detection of minimal ventricular volume) a specific threshold is marked in a separate signal using a unit impulse function, also termed a flag.

In order to locate the extremum, the part of the signal below the threshold is removed. The remaining part of the signal is differentiated and the result is cubed. Cubing actually reduces the noise level near the location of an extremum[11]. A smaller noise level at zero crossing point means more accurate positioning of the flag.

When a flag is generated at the position of a maximum, the flag is given a value of +1. Similarly, flags generated at the position of a minimum are given the value –1. The ECG flags receive the value +1 and are kept in a separate signal.

Only two consecutive values of the normalised signal are necessary in taking the decision to generate a flag. This means that the algorithm produces a delay of $1/F_s = 0.5$ ms. Consequently, the same flagging algorithm can be implemented in real time without difficulty.

The question of what is an appropriate threshold level is an important one because if defines the content of the flag signal. Ideally, the flag signal should be a periodic chain of flags, each indicating the position of a ventricular volume extremum. This could be true for perfectly clean TCG signals but in practice no signal is completely clean.

During the acquisition, a normal TCG signal will contain one or more localised perturbations due to spurious chest movements. These movements are due to overall body displacement, or frustrated ventilation reflex, or even digestion. The human body is a dynamic system and interrupting ventilation by holding completely still is not a natural position. In other words, localised perturbations are inherent to the TCG signal, even for highly motivated subjects. These disturbances are either major or minor. Minor disturbances are more frequent than major ones but do not distort the TCG signal excessively.

At the beginning of the acquisition, a few seconds will elapse between the moment when the subject is asked to stop breathing and the moment when his chest position becomes stable. Similarly, small convulsion might disturb the stability of his chest position at the end of the acquisition, when air is rarefied in his lungs. To limit these beginning and end effects, the TCG signal (of duration 30 s) was padded with zeros for two seconds at the beginning and the end of the

[11] Given the profile of the cube function, and also that the noise level is below the signal level, the noise amplitude is reduced more than the actual signal amplitude. This process was very important during an early stage of the research when the denoising filter was not as effective as it is now.

acquisition. The bandpass filter also contributes to the beginning effect by requiring an extra 0.5 to 1 s for its output to completely stabilise and track the input signal. The resulting TCG signal has an effective duration of 25 s.

During a disturbance, although the height of the maximum and minimum ventricular volumes can be severely altered, their waveforms are distorted but rarely completely obliterated and a local peak can still be observed. It results that the TCG ventricular volume extremum never has an entirely stable height but nevertheless is always visible.

Of course, the majority of flags will be generated at the location of a ventricular volume extremum where the TCG signal has a high amplitude. However, a ventricular volume extremum of lower amplitude should still be responsible for generating a flag, if it is above the threshold.

The consequence is that the signal level at an extremum cannot be the only recognition criterion since this information seems impossible to categorise: the flagging algorithm should have a low threshold to avoid missing flags.
By doing so, we will also increase the number of flags which were not generated at a ventricular volume extremum, and pollute the population of "good" flags. This is acceptable if we provide a way to sort good flags from bad flags.

12.4.1.5.2. Checking TCG flags validity

Among all the flags generated, some of them do not actually represent a ventricular volume extremum. The problem of finding and discarding erroneous flags is equivalent to tracking the cardiac periodicity within the flag signal. The first assumption here is that the flags of ventricular volume extrema are the only periodic flags, and form a chain in the flag signal.

The second assumption is that the heart rate will remain locally stable during the acquisition, in the sense that the heart period will not vary by more than 200 ms from one beat to another (unless the patient suffers from cardiac arrhythmia). These assumptions are verified in almost any situation and are therefore considered valid.

Given these two assumptions, the difference between a "clean" and a "dirty" flag signal should be a third flag signal containing only erroneous flags, i.e. flags that were not generated at an extremum of ventricular volume and that show no periodicity.

Four algorithms in total were written to discard erroneous flags. The first two presented below only work in offline mode. The remaining third and fourth operate in real time. They are presented later in section 12.2.1.4.2. The main structure of each algorithm is a loop scanning through the flag

signal. Each algorithm consists of a series of logical and numerical tests to help determine whether a particular flag is pointing at an extremum of the ventricular volume. The result of these tests is coded by the value of the flag being tested. All four algorithms use the same code, which is presented below:

Absolute value of flag	Associated significance
1.50	This flag is valid and the previous flag was higher or equal to 1.25.
1.25	This flag is valid but the previous flag was smaller than 1.25.
1.00	This flag has not been checked for validity yet.
0.75	This flag is probably valid.
0.50	This flag is probably not valid.
0.25	This flag is not valid. It will not take part in any other tests.

Table 2: Flag value significance

The difficulty in writing these algorithms was in making them process either clean or dirty signals with the same apparent quality of results. Very clean signals are very simple to process whereas very dirty signals seems to require the inclusion of lots of particular cases in the test structure, because there is no specific error pattern to them. The interactions inside the algorithm of multiple particular cases can sometimes produce unexpected results, which are difficult to understand.

The challenge was to generate the simplest possible structure, while permitting the successful treatment of the largest possible number of flag signals. In other words, it was attempted to generalise the particular cases.

As a golden rule, if the position of a flag cannot be explained in relation to the other flags, then no risk should be taken and this flag should be ignored. All algorithms were set to operate in between 2 and 28 seconds of the flag signal.

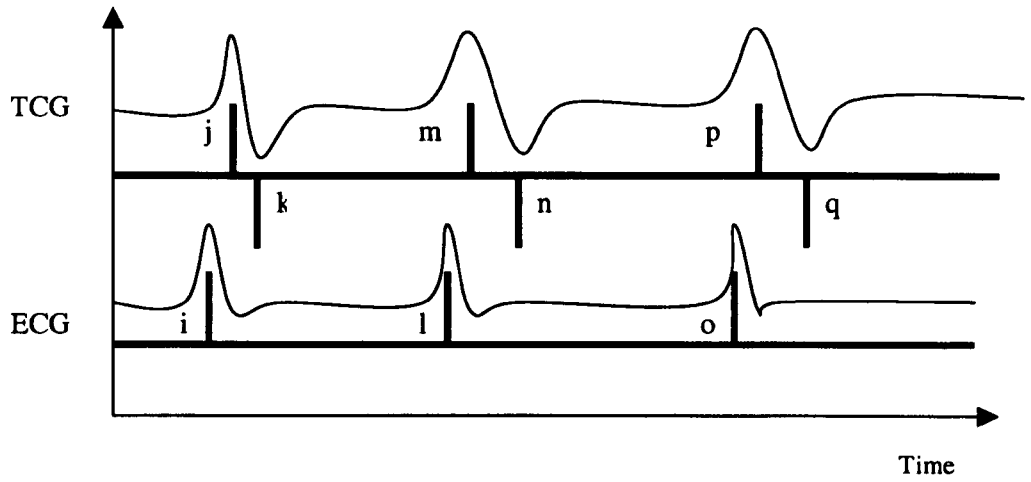

Figure 16: Designation of ECG and TCG flags necessary for explaining the validity checking algorithm. All flags have the same value (±1) because they have not been discriminated yet.

The first algorithm uses the ECG signal acquired in parallel with the TCG to spot erroneous flags. This algorithm has only one layer of tests. The tests consist in verifying that the interval between a particular positive flag (say "m") and the anterior ECG flag ("l" in this case) is similar to R2MAX (about 100 ms). If it is not, the positive flag is discarded. The process of discarding a positive flag is by giving it a "penal" value of +0.25. The penalised flag is still visible but is no longer taken into account by the algorithm. Negative flags are checked against R2MIN (about 200 to 250 ms) and erroneous flags are given the value –0.25.

If multiple positive flags can be found inside the same time region, only the latest one is kept. If multiple negative flags can be found inside the same time region, only the earliest one is kept. This ensures that the remaining pair of flags are the closest possible to each other, on both sides of the transition from maximum to minimum ventricular volume.

Note that checking the positive and the negative flags are two independent processes. The reason for this separation is to avoid penalising a positive/negative flag if the corresponding negative/positive flag was erroneous rather than discarding both of them. The algorithm is reproduced in appendix 3.

The second algorithm does not use ECG support and therefore has no knowledge of heart rate or R wave position before or during the analysis. The absence of a reference signal obviously makes the algorithm more complex than the first one, since more information has to be extracted from the TCG flag signal. The principle of the algorithm is based on the assumption proposed earlier that the heart rate is locally stable. In other words, in a chain of three valid flags[12], the duration of the

[12] Three is the minimal number of flags to observe a periodicity. A more complex offline algorithm could possibly use five or more flags to ensure that the local chain is part of the global chain of valid flags.

interval between the last two flags should be sensibly similar to the duration of the interval between the first two.

This second algorithm has two layers of tests. In the first layer, the algorithm will discard flags lying too close to each other. If two consecutive positive flags are closer than 200 ms, the earlier flag is discarded. If two consecutive negative flags are closer than 200 ms, the later flag is discarded. Discarded flags are given the penal value of ± 0.25, depending on their initial sign, and the algorithm will ignore them.

In the second layer, all the remaining flags are assumed to be probably wrong, they are given the value of ± 0.5. Then, if the present flag (either positive or negative) is found to lie approximately in the middle of two other flags of the same sign, the present flag is considered valid and is given the value of ± 1.25. If the first flag of the chain is also a valid flag, then the current flag is given the value of ± 1.50 instead, in order to reconfirm its validity. In this case, the algorithm will take the risk to render invalid (± 0.25) all the flags located between the current flag and the last flag in order to eradicate possible parallel interlaced chains of periodic flags. This is a recursive process in the sense that the second layer will have to deal in the future with its own decisions made in the present. Hopefully, the decision is a good one and the algorithm will speed up by avoiding unnecessary tests. As with every recursive process, there is also a risk of instability: a valid chain of flags may be invalidated by mistake. Or, an invalid chain of flags may be validated by mistake. In any case, this would result in a momentary incapacity to locate any valid flags.

Again, the treatment of positive and negative flags was kept separate to avoid creating a dependence of one on the other. Initially, this was not the case: the negative flags were tested first[13], and the result of the test (the duration of the interval between two valid negative flags, designated as MIN2MIN) was compared with its positive equivalent. While giving superb results for clean signals, actually helping to improve the accuracy of the tests on positive flags, this recursive dependence could also "crash" the algorithm when valid MIN2MIN intervals could not be generated. This is why separate tests were preferred for positive and negative flags. The algorithm is evaluated using Matlab as shown as in appendix 4.

Recursivity is very good at propagating information through the signal by influencing a decision taken in the present with the outcome of previous decisions. If successfully coded, the recursive information path contains an understanding of the signal's behaviour, which links different parts of the algorithm. However, recursivity is unaware of the validity of the information being propagated.

---

[13] Due to the large negative excursion of minimum ventricular volume, the negative threshold can be set very low. The result is that there are fewer negative flags than positive ones, hence fewer negative errors. Consequently, erroneous negative flags are easier to locate.

In this algorithm (and the others), the use of recursivity was avoided, except when it seemed to present minimal risks.

The results of the two offline algorithms are presented in the next two sections.

12.4.1.5.3. Occurrence of TCG flags

To measure the offline algorithms capability at locating valid and invalid flags, the occurrence of valid and invalid flags was measured by inspection for each subject, each type of flag (positive and negative), and each algorithm. The results are reported below in table 3. The group in the column labelled TCG MIN designates minimum ventricular volume, hence negative flags. Similarly, TCG MAX designates positive flags.

The label "found and valid" designates flags of level ± 1.25 or ±1.50 that were located in their expected time regions (100 ms for positive flags, 200 to 250 ms for negative flags). On the contrary, the label "found and invalid" designates flags of level ± 1.25 or ±1.50 that were not located in their expected time regions. Finally, the label "missed" designates flags that did not appear in their expected time region, either because they did not exist, or because they had a value smaller than ±1.25. The table visibly summarises the outcome of the algorithms for flag validity checks but, incidentally, the flagging, normalising and filtering processes are also involved in the quality of the results presented below.

COUNTING FLAGS OCCURRENCE WITH OFFLINE DETECTION							
Subject	ECG assisted detection	TCG MAX			TCG MIN		
		Found & Valid	Found & Invalid	Missed	Found & Valid	Found & Invalid	Missed
1	yes	26/27 (96.3%)	0	1 (3.7%)	2627 (96.3%)	0	1 (3.7%)
	no	25/27 (92.6%)	0	2 (7.4%)	25/27 (92.6%)	0	2 (7.4%)
2	yes	18/19 (94.7%)	0	1 (5.2%)	18/19 (94.7%)	0	1 (5.2%)
	no	14/19 (73.7%)	0	5 (26.3%)	15/19 (78.9%)	0	4 (21.0%)
3	yes	25/25 (100%)	0	0	24/25 (96.0%)	0	1 (4.0%)
	no	23/25 (92.0%)	0	2 (8.0%)	23/25 (92.0%)	0	2 (8.0%)
Average of all subjects	yes	97.0%	0.0%	3.0%	95.7%	0.0%	4.3%
	no	86.1%	0.0%	13.9%	87.8%	0.0%	12.2%

Table 3: Summary table of flags occurrence for all subjects during offline detection. Both ECG assisted and ECG independent algorithms are presented.

The results concerning the average of all subjects suggest that both algorithms are successful at separating valid and invalid flags. On average, during the 26 seconds of the analysis, 97% of positive flags and 95.7% of negative flags designated as valid by the ECG assisted algorithm were actually valid "for real". For the ECG independent algorithm, 86.1% of positive flags and 87.8% of

negative flags were accurately validated. This is about 10% less than with the ECG assistance but is still significantly high.

The maximal average difference between the percentage of positive and negative "found and valid" flags is only 1.7%, indicating that the algorithms are equally capable of detecting the maximum and the minimum ventricular volume.

More importantly, none of the algorithms has generated wrongly validated flags. Transposed to the context of medical imaging, this means that no image layer would have been inadvertently acquired.

12.4.1.5.4. Localisation of valid TCG flags

Among the flags that were labelled "found and valid", the statistical measurements of the R2MAX and R2MIN intervals are reported below in table 4, for each subject and each algorithm. Since these measurements are carried out on the TCG signal, the results below reflect the combined actions of the filtering and normalising processes.

MEASURING R2MAX AND R2MIN WITH OFFLINE DETECTION							
Subject	ECG assisted detection	R2MAX			R2MIN		
		MEAN (ms)	SD (ms)	SD/MEAN	MEAN (ms)	SD (ms)	SD/MEAN
1	yes	111.4	2.9	2.6%	206.3	7.0	3.4%
	no	112.0	2.7	2.4%	206.3	7.0	3.4%
2	yes	101.9	5.9	5.8%	254.1	15.6	6.2%
	no	101.8	6.4	6.3%	253.5	16.4	6.5%
3	yes	114.6	3.4	3.0%	204.9	8.1	4.0%
	no	114.5	3.5	3.1%	204.8	8.3	4.0%
Average of all subjects	yes	109.3	4.1	3.8%	221.7	10.3	4.5%
	no	109.4	4.2	3.9%	221.5	10.6	4.6%

Table 4: Summary table of R2MAX and R2MIN for "found and valid" flags.

These results show that, in average, the R2MAX interval has a smaller standard deviation than the R2MIN interval because it is more precisely located after the R wave.

The conclusion of this analysis of localisation, coupled with the previous analysis of occurrence, is that given the equivalent difficulty of localising valid positive flags and negative flags, positive flags have a more consistent localisation, and therefore form a better anchor.

12.4.1.5.5. Comparative cardiac period monitoring with ECG and TCG signals

To assess the performance and help localise the errors of the algorithms in checking flag validity, the instantaneous heart period is derived from two separate sources and the two period signals are compared. The first source is the R2R interval of the ECG signal, and the second source is the MAX2MAX interval (see figure 15) of the TCG signal. The purpose of the ECG period signal is to serve as a heart period reference. A MAX2MAX interval exists between two positive flags of level equal to or higher than 1.25. All three subjects are represented, for both algorithms. The period signal is represented over 26 seconds, from second 2 to second 28 but the bandpass filter and the flag validity checking algorithms will only start to stabilise at second 3. Consequently, the periods signals of both TCG and ECG between 2 and 3 seconds suffer from a transient starting effect and should not be taken into account.

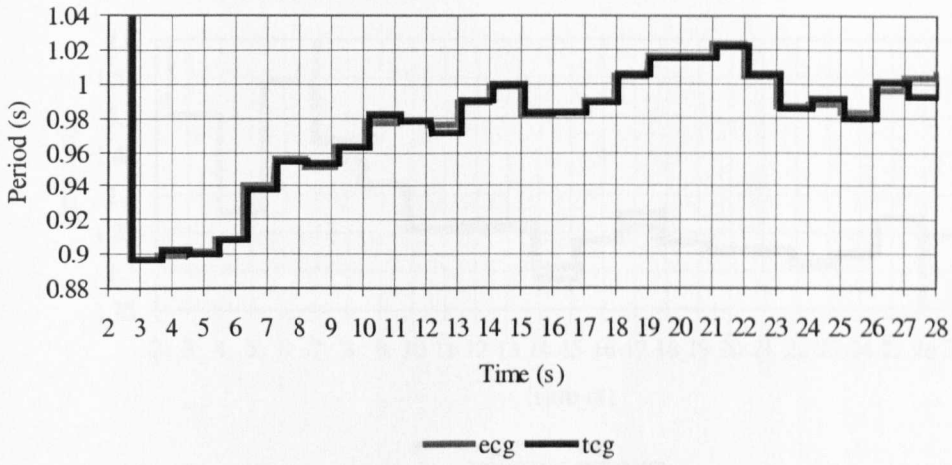

Figure 17: Cardiac period signal for subject 1, ECG assisted flag validity checking.

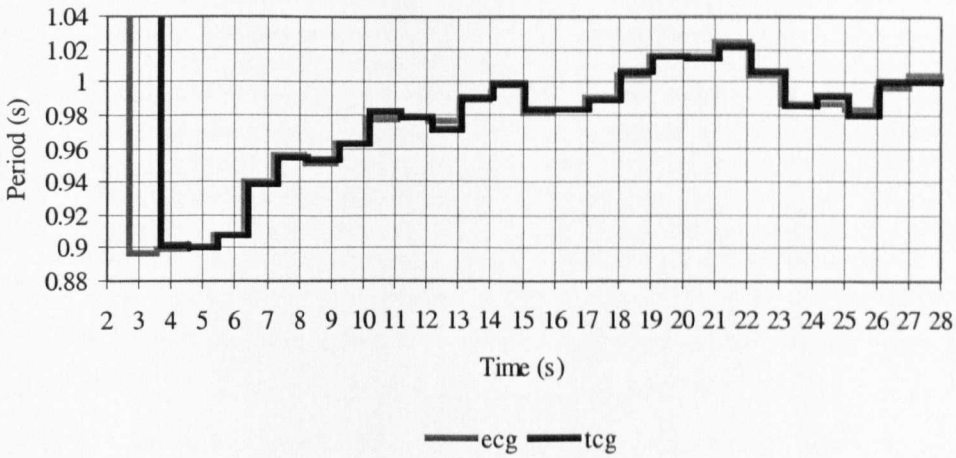

Figure 18: Cardiac period signal for subject 1, ECG independent flag validity checking.

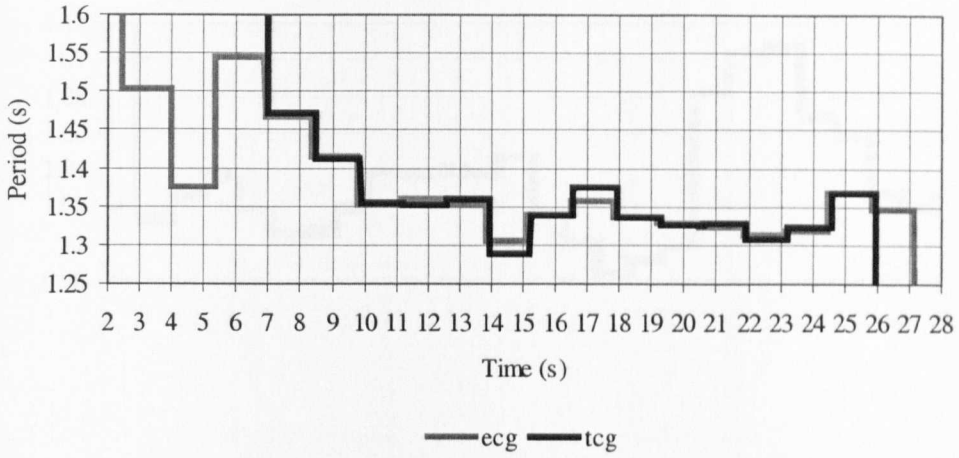

Figure 19: Cardiac period signal for subject 2, ECG assisted flag validity checking.

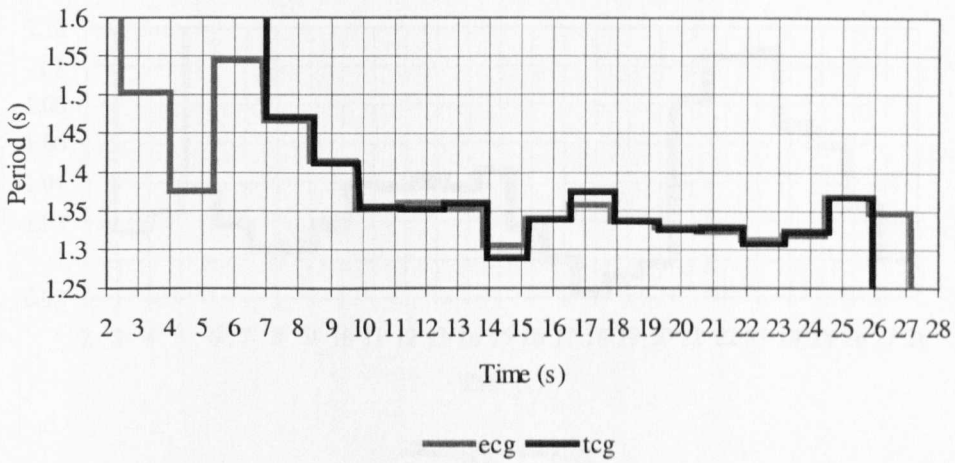

Figure 20: Cardiac period signal for subject 2, ECG independent flag validity checking.

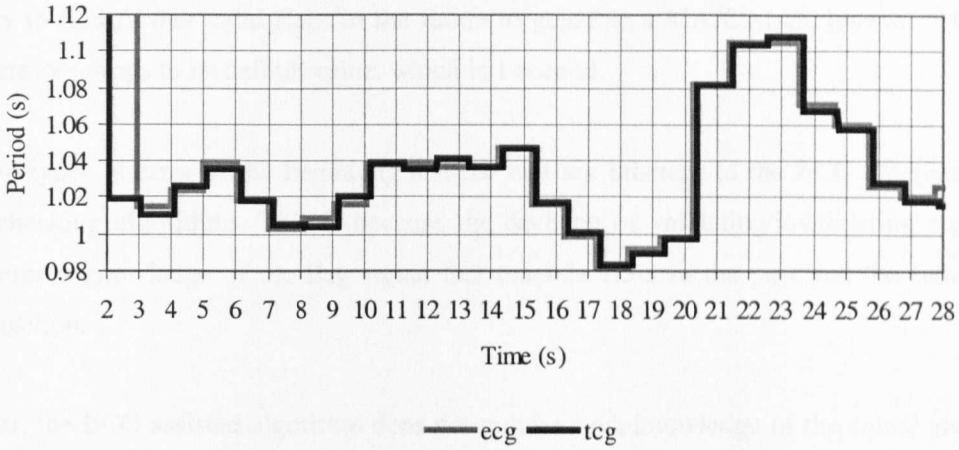

Figure 21: Cardiac period signal for subject 3, ECG assisted flag validity checking.

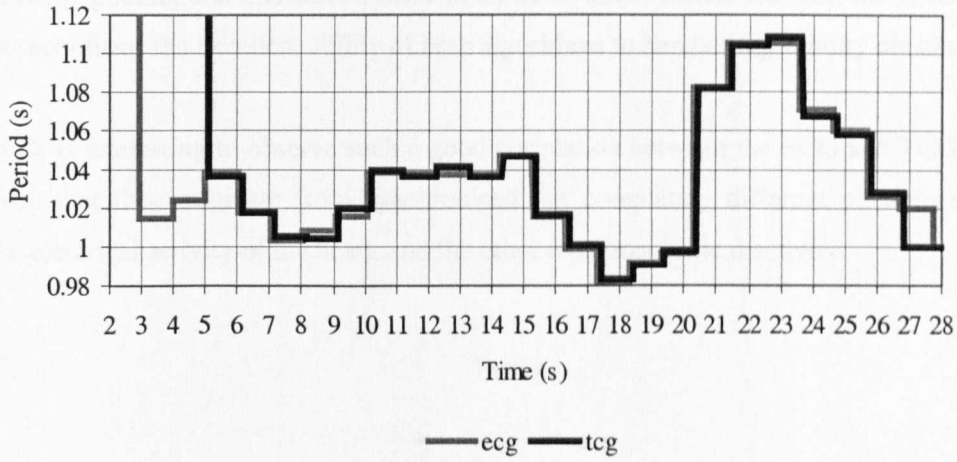

Figure 22: Cardiac period signal for subject 3, ECG independent flags validity checking.

By inspection of the graphs presented above, we can reconfirm that localisation of maximum ventricular volume is successful. Most errors are located at the beginning, between to 2 and 5 seconds, when the algorithm is attempting to "track the beat" of the flag signal. Other errors, not as frequent, are localised at the end, before second 28.

The beginning error correspond to a localised major overestimation of the period signal, which is due to missing valid flags. For example, if a single flag is missing, or if it doesn't have a big enough value to be considered valid, then the measured period is not the heart period but approximately twice as much, hence a large localised increase in the period signal. This happens at the beginning because there is no previous flag.

The end error corresponds to an abrupt fall of the period signal to 1 second. This is due to the incapacity to find further valid flags in the future to generate a MAX2MAX interval. The period signal therefore drops to its default value, which is 1 second.

These two types of error at the beginning and the end are inherent to the ECG independent flag validity checking algorithms. This is because the decision of validating/invalidating a particular flag requires a knowledge of the flag signal that extends 1.5 s in the past and the future of the current position.

In contrast, the ECG assisted algorithm does not require such knowledge of the signal and third is why the period signal derived in this way has a smaller dead time than its ECG independent counterpart.

Apart from the beginning and end errors, there is no other major difference with the ECG period signal. This reconfirms the excellent ability of both algorithms to handle flag validity checks.

Incidentally, it is interesting to observe such a good correlation between the ECG and TCG period signals, given that they originate from synchronised but completely different phenomena: one relates to the electrical activity of the heart, and the other to its mechanical activity.

12.4.2. Real-time study of TCG signals

The modes of real time and offline analysis of the TCG signal are similar: in order to trigger the scanner, a suitable anchor must be located in the signal. For this matter, the signal is inspected and its statistics are studied in detail. Subsequently, an anchor is proposed and a specific algorithm is written to attempt to detect the anchor, in real time in this case. The results of this algorithm are assessed by comparison with a parallel ECG recording, and from this it is concluded whether or not the anchor detection is consistent enough to serve as a scanner trigger.

Both offline and real time analysis use the same category of processes, i.e. filtering, normalising, flagging and flag validity checking. Filtering and normalising are required to permit visualisation and inspection of the TCG signal, whereas flagging and flag validity checks are required after filtering and normalising to permit anchor detection. The filtering and flag validity checking processes are of particular importance in the conversion from offline to real time analysis.

Real time analysis of the TCG signal is a more complex task than the offline analysis since a process never has access to the future of the signal. In other words, the amount of information available at run time is lower and everything required to produce successful anchor detections must be extracted and processed from the past of the TCG signal. This must be done as fast as possible to limit detection delays. The implementation of this analysis was performed exclusively on Matlab in view of a later implementation on a true real time DSP system.

The forthcoming subsections have the same structure as section 12.4.1 (offline study of TCG signals).

12.4.2.1. Filtering

The process of real time filtering is always problematic. Filters, either FIR or IIR, have a propagation delay, or phase delay, and therefore create a lag or a lead between the input and the output signals. The usual assumption in this case is that it is acceptable for a filter to exhibit a phase delay, as long as the delay is short, known and invariable. With their inherent linear phase, FIR filters produce a constant and easily predictable phase delay.

For this reason, the first attempt to create a real time band-pass filter for TCG analysis was by using a combination of two moving average filters. The interest in using moving averages was that the convolution process was limited to a sum, instead of a sum of products, therefore simplifying and accelerating the filtering process[1]. A first moving average would denoise the signal. Then, the denoised signal was branched to a second moving average. The output of this second moving

[1] This reduction of the computational burden would have been of particular importance if the filter implementation had been transposed to a smaller system, such as a DSP chip or even a microcontroller.

average was delayed by half the duration of the sum, and then subtracted from the denoised signal. This created the detrending effect.

The filter gave satisfactory results in terms of signal to noise ratio, but at the expense of long averaging times. In total, up to 500 ms of signal history was required to produce a satisfactory output signal, therefore delaying the output signal by 250 ms. Clearly, such a delay could not qualify as a real time process[2] and reducing the averaging duration only decreased the signal to noise ratio.
It was therefore decided to use an IIR filter, despite its non-linear phase response (and consequently its varying phase delay). The assumption made was that it is acceptable to use an IIR filter, providing that the phase response in the pass band is as linear as possible. Typically, Bessel filters offer the smallest deviation from linearity in the pass band of their phase response, ensuring a minimum variation of the phase delay.

Such a Bessel filter was realised in Matlab. Surprisingly, Matlab 5.3 did not provide any function to generate directly the digital coefficients of a Bessel filter. (In contrast, other filter prototype functions, such as "butter" (Butterworth filter) and "cheby" (Chebyshev filter) can immediately produce digital coefficients.) The Matlab function "besselap" in the signal processing toolbox permitted the generation of transfer functions of low and high pass filters[3]. To obtain digital Bessel filters, the two transfer functions were then converted from the Laplace domain to the Z domain using a bilinear transform. Finally, the digital pass-band Bessel coefficients were generated by taking the polynomial product of the coefficients of each filter. This method led to stable filter coefficients (the impulse response of the passband filter is converging to zero).

The cut off frequencies of the band pass filter are the same as the offline filter, i.e. F_{high} = 20 Hz, with order 2, and F_{low} = 1 Hz, with order 4. The order of the low pass side was diminished compared to the offline filter so as to limit the phase delay in the passband.
The frequency response of this filter, and a comparative Butterworth filter designed with the same characteristics, are represented below in figure 25.

[2] If the filter produces a constant delay of 250 ms, no triggering can be done for 250 ms after the detection of the anchor. This might be acceptable for anchors located closely to the R wave, but generally speaking, a reduction of the filter delay is desirable.
[3] The analog prototype function "besselap" could also generate transfer function of band pass filters, with the limitation of having both slopes of the same order. This did not fit our needs and this is why it was decided to combine two separate Bessel filters in one.

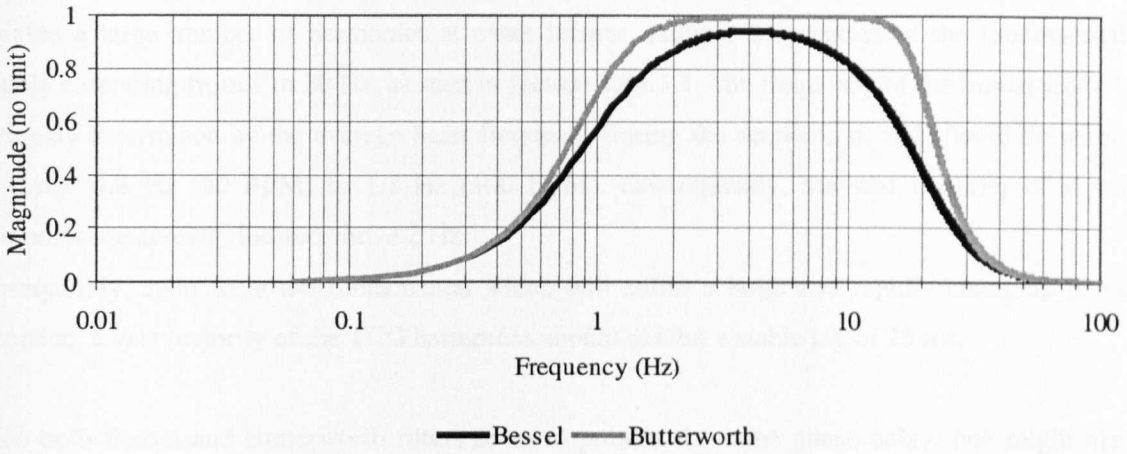

Figure 25: Comparative frequency responses of Bessel and Butterworth filters, designed with the same characteristics. The passband is 1 to 20 Hz. As expected, the Butterworth filter has a steeper roll off than the Bessel filter.

Although Bessel filters are known for their maximally flat phase delay in the passband, this property does not seem to apply very well to high pass filters, especially when designed with a very low cut-off frequency such as 1 Hz.

In fact, the phase delay curves of Bessel and Butterworth filters, designed with the same characteristics, are no different from each other. In other words, the Bessel filter apparently does not show any reduction of phase delay, compared to a standard Butterworth filter. See figure 26 below.

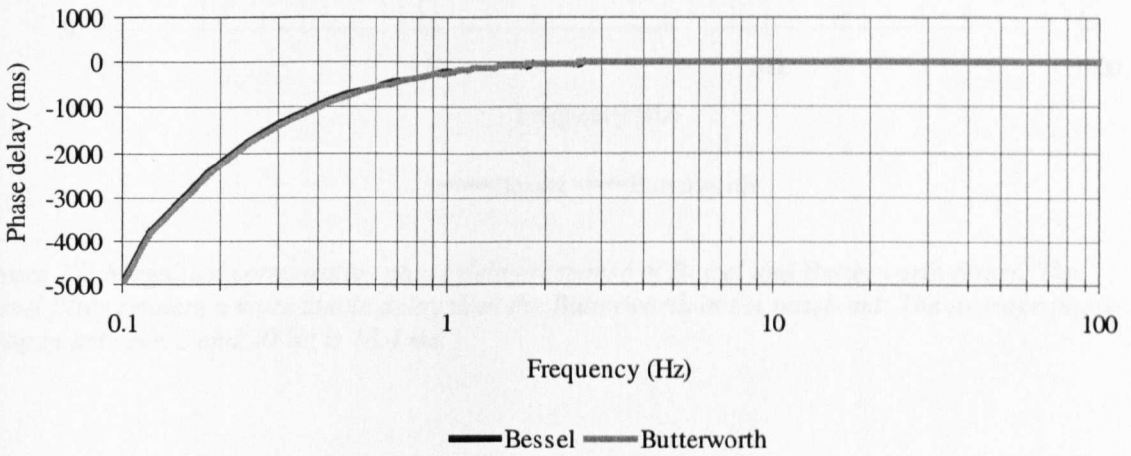

Figure 26: Comparative phase delay response of Bessel and Butterworth filters, designed with the same characteristics. The passband is 1 to 20 Hz. A positive delay indicates a lag between the filtered and the original harmonics. The Butterworth curve is superimposed on the Bessel curve, indicating that both filters have similar phase delay responses.

In the graph above, we can observe that the phase delay response of the Bessel filter has a catastrophic value of –750 ms at 0.5 Hz and –250 ms at 1 Hz. Beyond 2 Hz, the curve is then rapidly stabilizes to 25 ms. Given this enormous variation, one might wonder what kind of consequences this phase delay response will have on the filtered signal.

The spectral analysis of the TCG signal, processed offline with a zero phase Butterworth filter, revealed a large number of harmonics at exact integer multiple frequencies of the fundamental, roughly extending from 1 to 20 Hz, as seen in section 12.4.1.3. The frequency of the fundamental is obviously determined by the average heart frequency during the acquisition, and should lie within the range 0.8 Hz (50 BPM) to 1.7 Hz (100 BPM). Consequently, the vast majority of signal harmonics are actually located above 2 Hz.
Subsequently, apart from the fundamental which will suffer a large and rapidly changing phase distortion, a vast majority of the TCG harmonics should exhibit a stable lag of 25 ms.

Since both Bessel and Butterworth filters seem to present the same phase delay, one might also wonder if a Bessel filter should be preferred to a Butterworth filter, or any other filter. By magnifying the previous graph, we can see that the Bessel filter exhibits a more stable phase delay than the Butterworth filter in the region from 2 to 20 Hz, where the harmonics of the TCG are localised. For this reason, it is finally concluded that the Bessel filter is preferable to the Butterworth. The magnified phase delay responses of both filters are represented below.

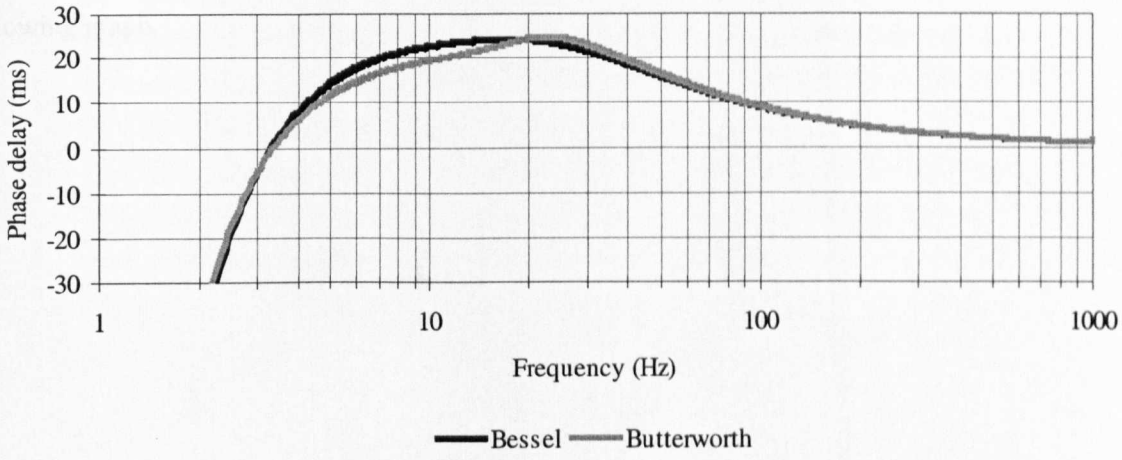

Figure 27: Magnified comparative phase delay response of Bessel and Butterworth filters. The Bessel filter renders a more stable delay than the Butterworth in the passband. The average phase delay in between 2 and 20 Hz is 18.4 ms.

12.4.2.2. Normalising

The real time and offline normalising algorithms are similar. However, the real time version must use a causal moving average to calculate the moving standard deviation. Given that this moving average has a duration of 1 second, the standard deviation has a constant phase delay of 500 ms with respect to the un-normalised signal[4]. This delay does not directly affect the normalised signal, but the way this signal is created. As a consequence, the real time normalisation process has a slower reaction speed and might generate erroneous gain corrections if the variations of the input signal are too rapid. However, on a global scale, the algorithm's capability remains unchanged.

12.4.2.3. Analysis of waveform

TCG signals filtered and normalised in real time are now represented using the previous system of ECG synchronised signal slices. The object of this representation is once again to localise anchors in the signals of the three subjects. The average and the standard deviation signals are also generated but the stability estimator is not required here. Such an estimation of the stability is irrelevant because the TCG waveform is distorted by the phase delay, as we will see on the following graph.

[4] A symmetric FIR filter has a phase delay of half the duration of the FIR vector, see proof in appendix 2.

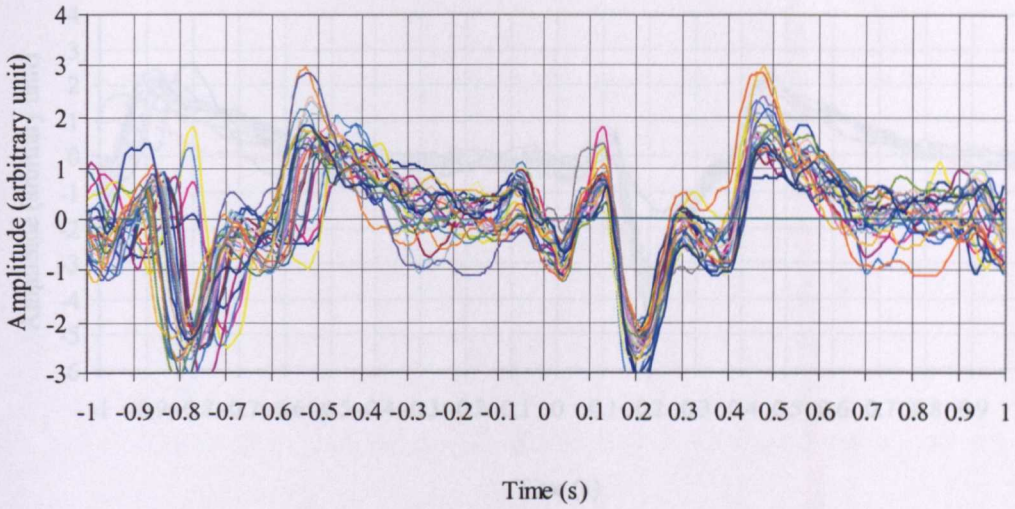

Figure 28:TCG signal slices of subject 1. The maximum and minimum ventricular volumes can be recognised at 100 and 200 ms. However, the maximum ventricular volume is considerably lower than during offline analysis. The minimal ventricular volume has the highest amplitude of all events in the signal. The real time signal peak-to-peak amplitude remains constant in comparison with the offline signal. A major peak appears at 450 ms. This peak was originally an inflexion point in the filtered TCG signal.

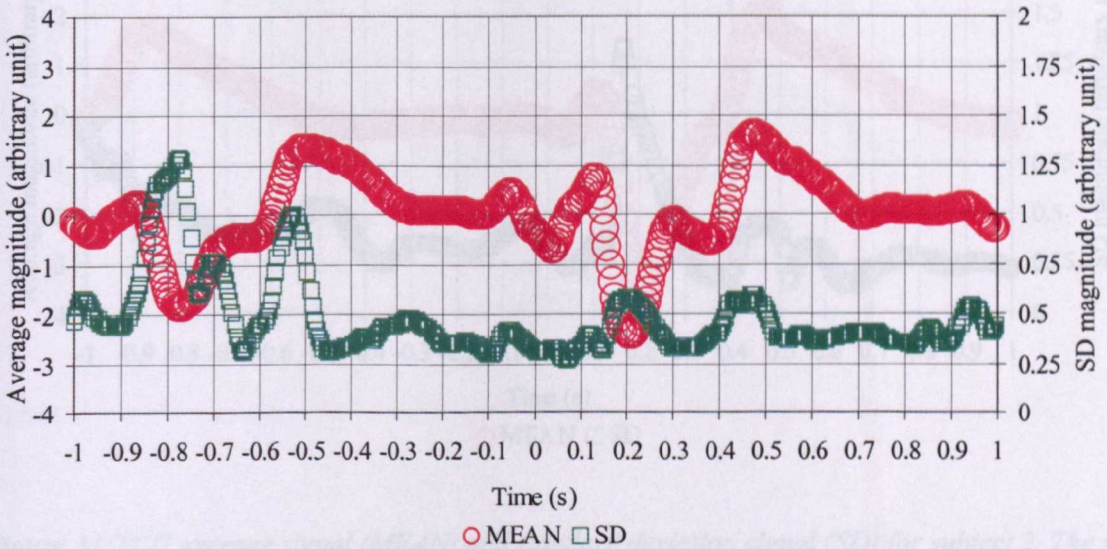

Figure 29: TCG average signal (MEAN) and standard deviation signal (SD) for subject 1. The overall SD signal has a level comparable to its offline equivalent. This indicates that the real time filter is consistent in the way the TCG signal is distorted.

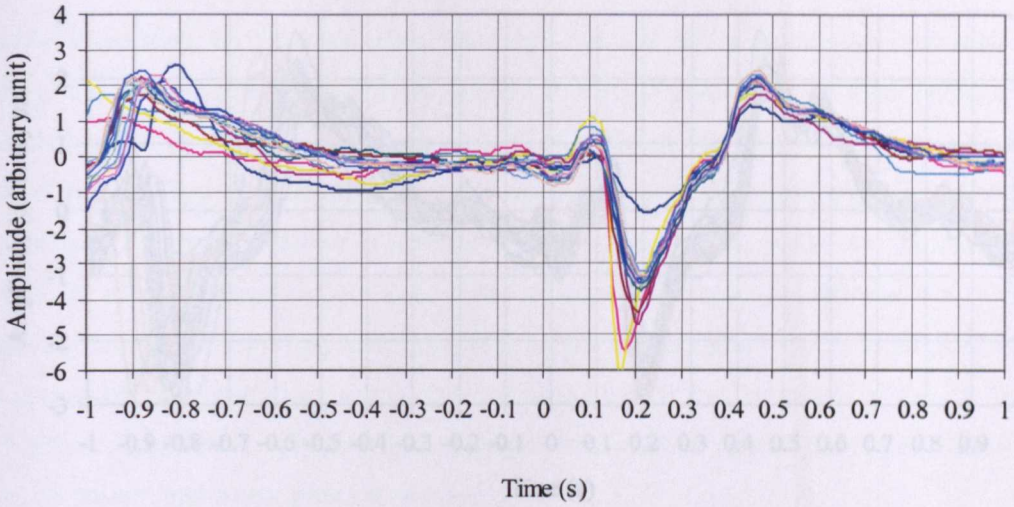

Figure 30: TCG signal slices of subject 2. The minimum ventricular volume is present at 200 ms and, like subject 1, seems exaggerated. This is a feature which would make it very easy to detect algorithmically. In contrast, the maximum ventricular volume at 100 ms seems truncated. Again, a major peak appears at 450 ms where no peak used to be in the offline filtered TCG signal.

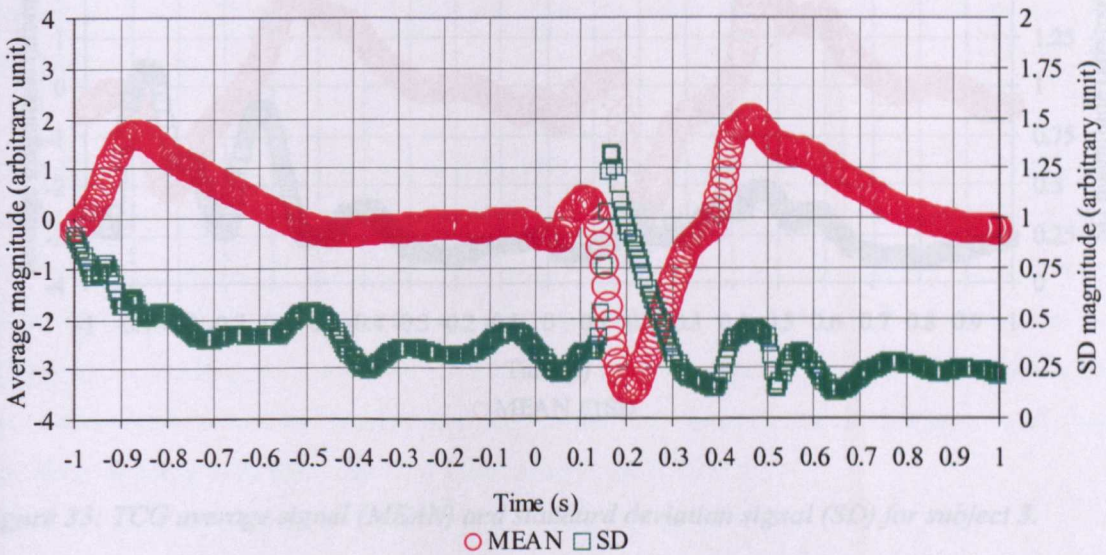

Figure 31:TCG average signal (MEAN) and standard deviation signal (SD) for subject 2. The real time SD signal has a different profile than its offline equivalent. This suggests that the distribution of regions with good repeatability has been altered.

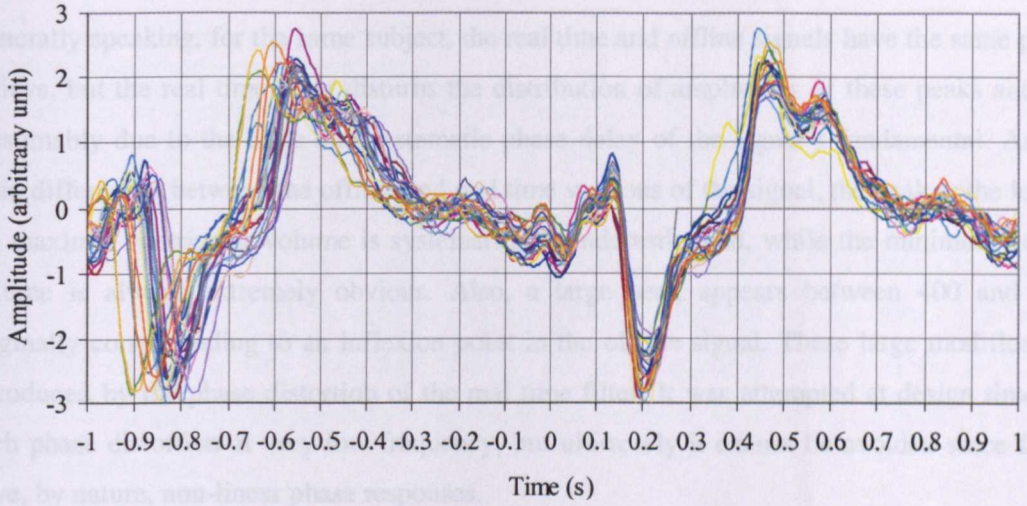

Figure 32:TCG signal slices of subject 3. This subject's signal presents the same characteristics as the first two, with the difference that the major peak at 450 ms has a higher amplitude.

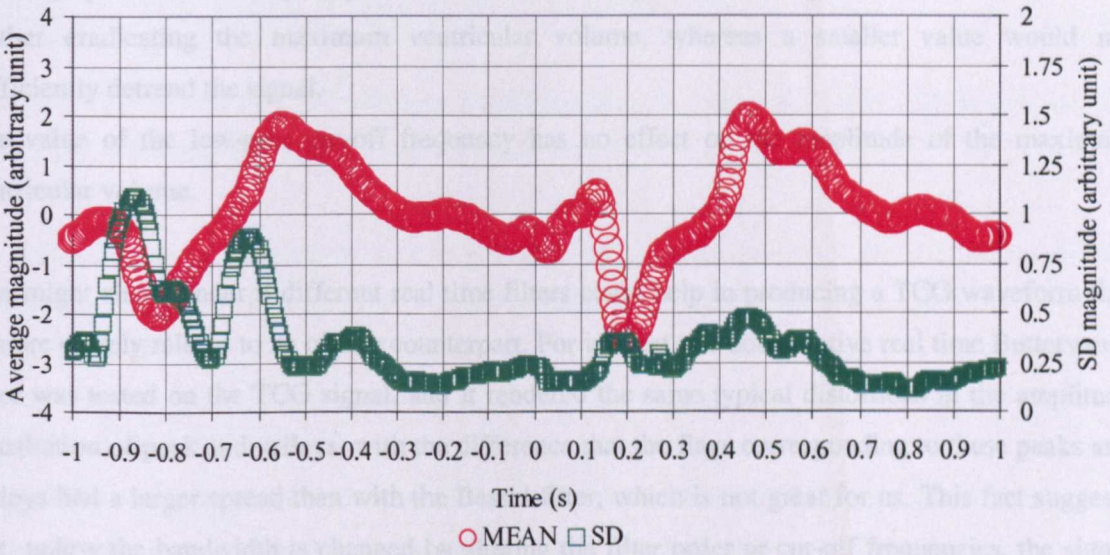

Figure 33: TCG average signal (MEAN) and standard deviation signal (SD) for subject 3.

Generally speaking, for the same subject, the real time and offline signals have the same peaks and valleys, but the real time filter disturbs the distribution of amplitudes of these peaks and valleys, presumably due to the large and systematic phase delay of the signal's fundamental. Among the main differences between the offline and real time versions of the signal, the peak at the location of the maximal ventricular volume is systematically underestimated, while the minimum ventricular volume is always extremely obvious. Also, a large peak appears between 400 and 500 ms, originally corresponding to an inflexion point in the offline signal. These large modifications are introduced by the phase distortion of the real time filter. It was attempted at design time to limit such phase distortion at very low frequency, but ultimately it cannot be avoided since IIR filters have, by nature, non-linear phase responses.

One might wonder if altering the cut-off frequencies of the band-pass filter could have a benefit effect on the TCG waveform.
The high-pass cut-off frequency was chosen as a compromise: A higher value would result in further eradicating the maximum ventricular volume, whereas a smaller value would not sufficiently detrend the signal.
The value of the low-pass cut-off frequency has no effect on the amplitude of the maximum ventricular volume.

One might also wonder if different real time filters could help in producing a TCG waveform that is more closely related to its offline counterpart. For interest, the comparative real time Butterworth filter was tested on the TCG signal, and it rendered the same typical distortions in the amplitude distribution of peak and valleys, with the difference that the flags corresponding to these peaks and valleys had a larger spread than with the Bessel filter, which is not great for us. This fact suggests that, unless the bandwidth is changed by altering the filter order or cut-off frequencies, the signal waveform is going to remain similar to what we have seen above.

Finally, the normalisation process has no responsibility in producing phase distortions: waveforms checked with and without normalisation have a similar aspect, despite spurious localised differences.

12.4.2.4. Algorithmic anchor detection

The problem of proposing potential anchors remains to be answered. Evidently, the minimum ventricular volume is a strong candidate since all subjects exhibit the same consistent valleys at 200 ms. Alternatively, another consistent event is the large peak located at 450 ms. However, it is far away from the R wave and, more importantly, it does not correspond to a specifically recognised event of the cardiac cycle. It probably is an artefact generated by the filter. The maximum ventricular volume is very consistent. Despite its underestimated level, it could be a potential anchor if the negative flag generation algorithm was modified so as to detect peaks in a particular band of amplitudes.

To be consistent with the offline analysis, it was decided to study the possibility of defining the maximum and minimum ventricular volume as anchors.

12.4.2.4.1. Flagging

The real time and offline processes to generate negative flags are exactly similar, except the adjustment of the negative threshold level, which was set to –1.5.

In contrast, the generation of positive flags had to be modified to search peaks in the amplitude band where the maximum ventricular volume is located, i.e. between 0 and 1.2.

12.4.2.4.2. Flag validity checking

To produce a real time estimation of whether flags generated by the flagging algorithm are correct or incorrectly located, two algorithms were written. The first one works in combination with a parallel ECG recording and the second is completely self-sufficient.

The first algorithm is simply based on its offline counterpart. Assuming that the ECG system produces a signal of constant, precisely measured and minimal phase delay, the algorithm will check the duration of the interval between a negative TCG flag and the preceding ECG flag. If this duration is similar to 200 ms, the negative flag is validated. If not, it is discarded. Additionally, if two negative flags are too close to each other, only the earliest is kept.

Having determined the position of a valid negative flag, the algorithm will scan the flag signal in the time region where the positive flag corresponding to the maximum ventricular volume was supposedly located. If is found, it is immediately validated. If multiple positive flags are found, only the latest is validated.

The second algorithm is based on the offline, ECG independent algorithm previously presented. The principle of locating chains of three, equally spaced, flags remains but the algorithm was modified to search for two flags in the past of the flag signal. Given the large and distinctive amplitude of the minimal ventricular volume, the population of corresponding negative flags contained less erroneous flags than the population of positive flags. For this reason, the negative flags were processed first. Once a negative flag was validated, the algorithm scanned back in time to attempt locating the corresponding positive flag. If multiple positive flags are found, only the latest is validated.

12.4.2.4.3. Occurrence of TCG flags

The results of the flagging and flag validity checking algorithm are summarised below in table 5. This table was constructed exactly in the same manner as its offline counterpart.

COUNTING FLAG OCCURRENCE WITH REAL TIME DETECTION							
Subject	ECG assisted detection	TCG MAX			TCG MIN		
		Found & Valid	Found & Invalid	Missed	Found & Valid	Found & Invalid	Missed
1	yes	22/27 (81.5%)	0	5 (18.5%)	25/27 (92.6%)	0	2 (7.4%)
	no	20/27 (74.0%)	0	7 (25.9%)	22/27 (81.5%)	0	5 (18.5%)
2	yes	16/19 (84.2%)	0	3 (15.8%)	17/19 (89.5%)	0	2 (10.5%)
	no	13/19 (68.4%)	0	6 (31.6%)	13/19 (68.4%)	0	6 (31.5%)
3	yes	24/25 (96.0%)	0	1 (4.0%)	24/25 (96.0%)	0	1 (4.0%)
	no	21/25 (84.0%)	0	4 (16.0%)	21/25 (84.0%)	0	4 (16.0%)
Average of all subjects	yes	87.2%	0.0%	12.8%	92.7%	0.0%	7.3%
	no	75.5%	0.0%	24.5%	78.0%	0.0%	22.0%

Table 5. Summary table of flags occurrence for all subjects during real time detection.

The results concerning the average of three subjects indicate that both algorithms are successful at detecting and validating/invalidating flags.

On average, during the ECG assisted detection, 92.7% of negative flags designated as "found and valid" were indeed valid flags. In comparison, 87.2% of "found and valid " positive flags were valid in the reality. For ECG independent detection, 78% of negative flags, and 75.5% of positive flags labelled "found and valid" were acknowledged as such.

These results indicate that, wherever the negative flags are, their detection is more efficient that the detection of positive flags. These results also indicate that ECG assisted detection is more efficient than ECG independent detection.

In comparison, real time detection yields slightly less accurate results than offline detection. Finally, like the offline algorithms, the real time algorithms did not validate any wrongly located flags.

The average location and the dispersion of positive and negative "found and valid" flags are discussed below.

12.4.2.4.4. Localisation of TCG flags

For the flags that were labelled "found and valid", the average and the standard deviation of the duration of R2MAX and R2MIN intervals were measured. The results are reported in table 6 below.

MEASURING R2MAX AND R2MIN WITH REAL TIME DETECTION							
Subject	ECG assisted detection	R2MAX			R2MIN		
		MEAN (ms)	SD (ms)	SD/MEAN	MEAN (ms)	SD (ms)	SD/MEAN
1	yes	131.7	3.8	2.9%	208.7	9.1	4.4%
	no	131.8	4.0	3.0%	209.5	9.4	4.5%
2	yes	109.9	5.5	5.0%	210.1	14.6	6.9%
	no	110.3	5.9	5.3%	213.4	5.0	2.3%
3	yes	135.0	3.3	2.5%	206.2	7.3	3.5%
	no	135.2	3.5	2.6%	205.7	7.7	3.7%
Average of all subjects	yes	125.5	4.2	3.5%	208.3	10.3	5.0%
	no	125.7	4.5	3.7%	209.5	7.4	3.5%
Difference with offline detection	yes	16.2	0.1	-0.3%	-13.4	0.1	0.4%
	no	16.3	0.3	-0.3%	-12.0	-3.2	-1.1%

Table 6. Summary table of R2MAX and R2MIN measurements with real time detection.

The average results show that, in real time, the localisation of positive flags is consistent, with or without knowledge of the ECG during detection. The average ± SD values of R2MAX for all subjects are 125.5 ± 4.2 ms with ECG assistance, and 125.7 ± 4.5 ms without ECG assistance.

Compared with offline detection, the real time average ± SD values of R2MAX show a difference of 16.2 ± 0.1 ms (with ECG) and 16.3 ± 0.3 ms (without ECG). These extremely small SD values strongly indicate that the average bias of 16 ms in the positions of positive flags measured offline and in real time is almost completely systematic and can very safely be approximated as a constant. In other words, the real time and offline positive maximum ventricular volume are equally dispersed, but around two centres located 16 ms away from each other.

The negative flags also seem to show a tendency to remain equally dispersed, either offline or in real time. In this case, the difference between the position of real time and offline averages is somewhere between –12 and –13 ms. This is not as precise as the positive flags and shows that the bias between offline and real time detection is not completely stable.

Coupling the analysis of occurrence and location shows that the positive flags, despite their somehow less efficient detection compared to negative flags, lead to a precise and accurate knowledge (once the bias has been subtracted) of the maximum ventricular volume. The bias of 16 ms is smaller than the expected value of 18.4 ms but not very different[5].

It was decided that the maximum ventricular volume is a good anchor. This choice has the advantage of applying also with offline detection.

Having found an anchor, it can be concluded that TCG detection in real time is a successful process.

12.4.2.4.5. Comparative cardiac period monitoring with ECG and TCG signals

To localise the occurrence of errors in the detection of positive flags, the instantaneous cardiac period signals of all subjects, with or without ECG assistance, are now represented. For general comments regarding the generation and the interpretation of these curves, refer to the offline analysis.

[5] The expected value of 18.4 ms was obtained by averaging the phase delay between 2 and 20 Hz.

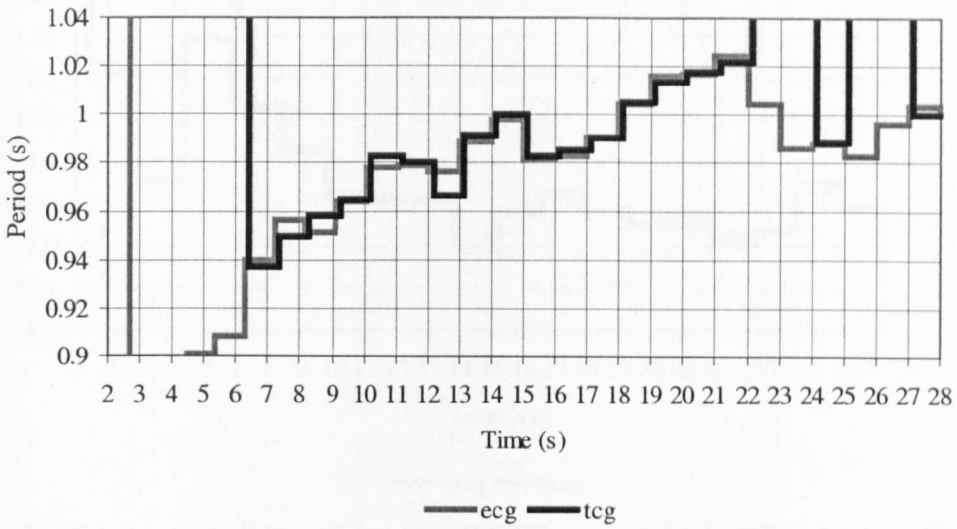

Figure 34: Cardiac period signal for subject 1, ECG assisted flag validity checking. It takes three seconds for the algorithm to start validating positive flags. The flag at 26 s is missed, resulting in a localised increase in the period.

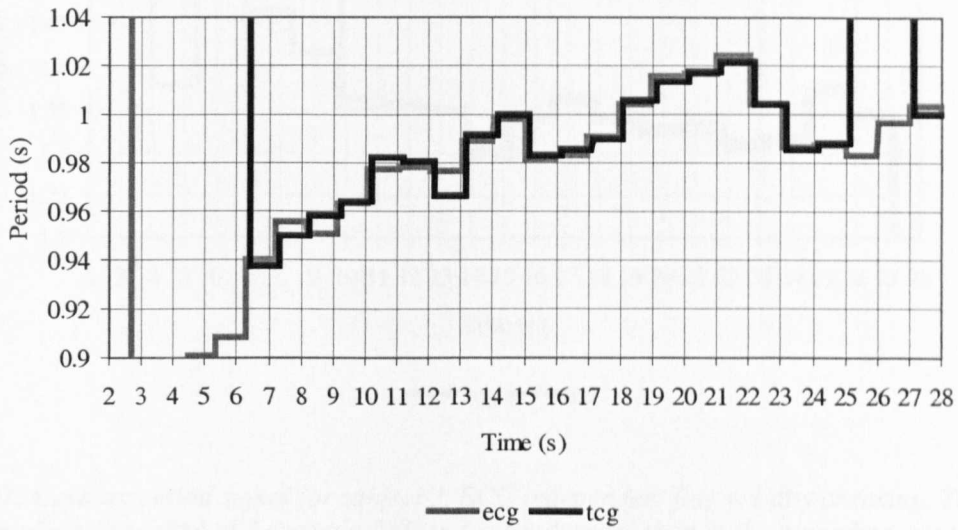

Figure 35:Cardiac period signal for subject 1, ECG independent flag validity checking. Same remarks as for the preceding graph. The algorithm also misses a flag at 23 s.

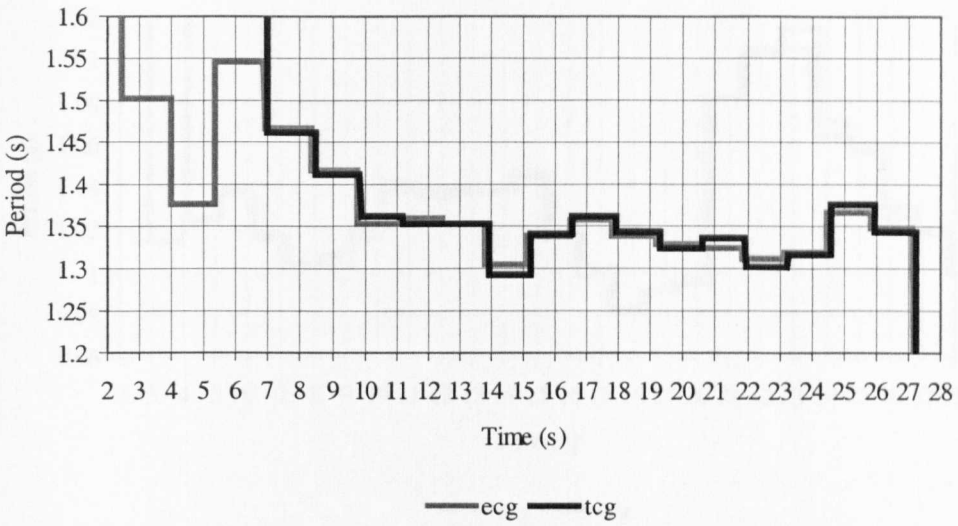

Figure 36: Cardiac period signal for subject 2, ECG assisted flag validity checking. The algorithm has a dead time of 4 seconds before generating flag validation. No flag is missed after this time.

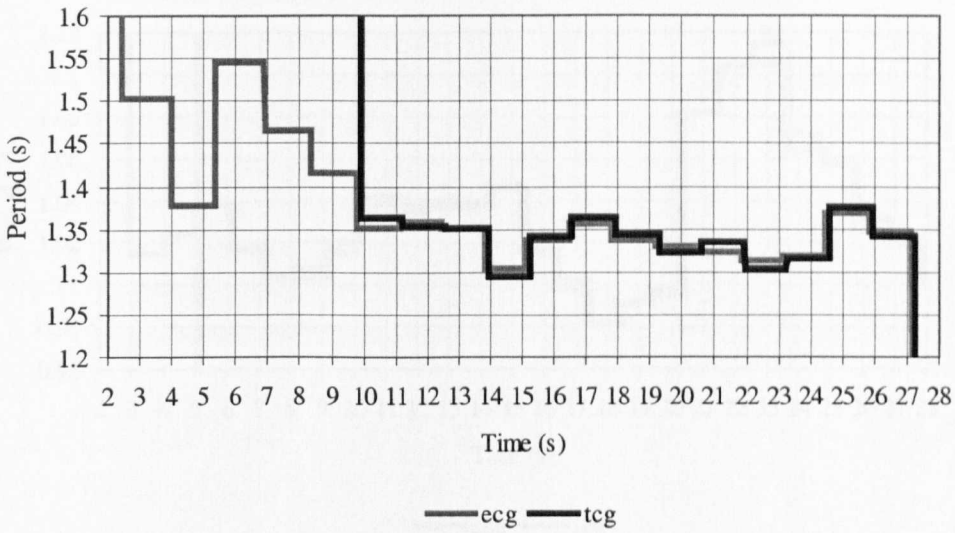

Figure 37: Cardiac period signal for subject 2, ECG independent flag validity checking. The algorithm ahs a dead time of 7 seconds. This is 3 seconds more than in the preceding graph. Beyond this point, no flags are missed.

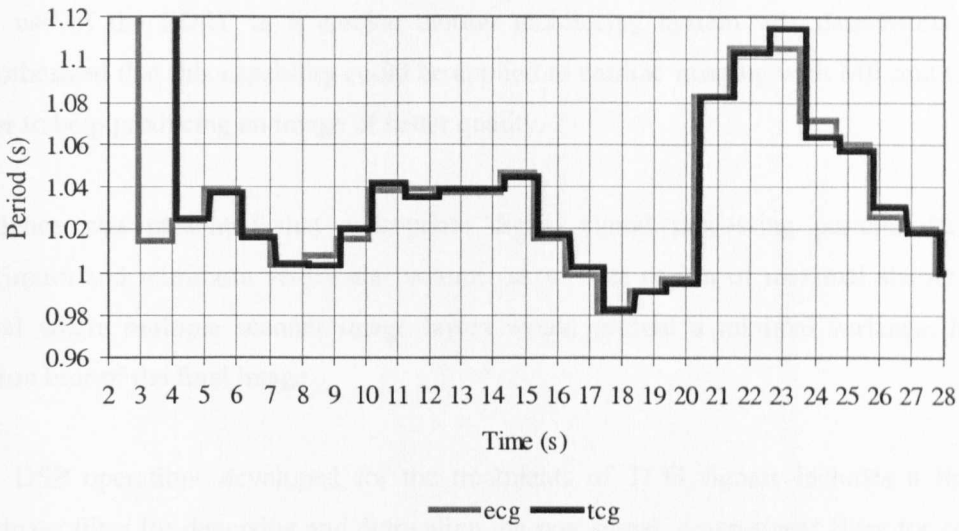

Figure 38:Cardiac period signal for subject 3, ECG assisted flag validity checking. The algorithm exhibits a dead time of 1 s. The onward TCG period signal remains close the ECG period signal.

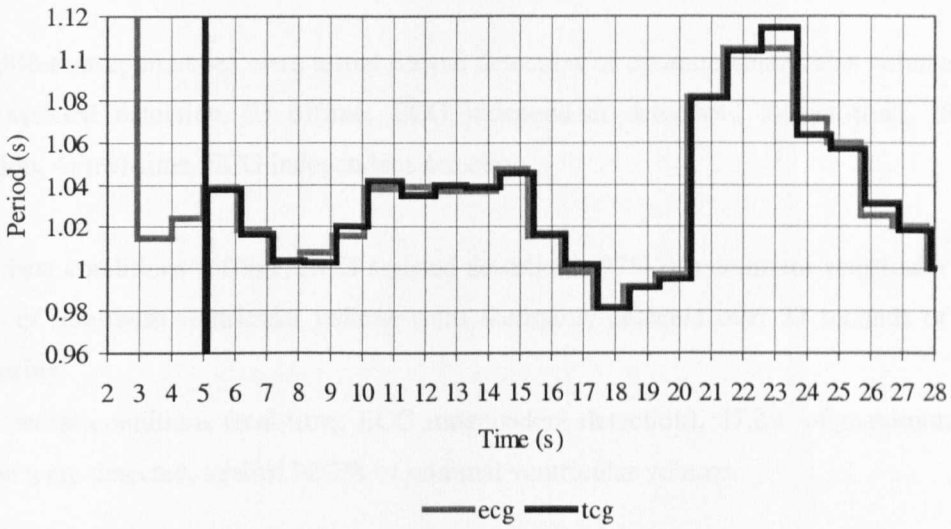

Figure 39: Cardiac period signal for subject 3, ECG independent flag validity checking. It takes two seconds for the algorithm to start validating flags. This is one second longer than the preceding period signal. This difference is attributed to the absence of ECG assistance during the detection.

In general, the ECG assisted algorithms show a smaller dead time than ECG independent algorithms. Once this dead time is over, both algorithms become almost equally efficient at validating/invalidating flags.

The duration of the dead time seems to decrease with a stable cardiac period, and increase with a large cardiac period. This is a logical result, with respect to the algorithm's implementation.

Finally, the dead time is slightly higher during real time analysis than during offline analysis.

12.5. Conclusion

The use of the FORP as a cardiac motion monitoring system was demonstrated and it was hypothesised that this capability could be applied to cardiac imaging with MR and CT scanners in order to help producing an image of better quality.

Evidence was presented that appropriate digital signal processing permits the detection of maximum and minimum ventricular volume, as well as region of maximal stability in the TCG signal where multiple scanner image layers would present a minimal variance, hence limiting motion blur of the final image.

The DSP operations developed for the treatments of TCG signals includes a linear recursive bandpass filter for denoising and detrending the raw signal, a non-linear filter for conforming the filtered signal into a specific amplitude range, an algorithm for detecting maximal and minimal ventricular volume and a signal stability estimation function for locating the most flat and repeatable regions of the TCG signal.

Four different approaches were tested for the detection of extrema ventricular volume: 1) Offline, ECG assisted detection, 2) offline, ECG independent detection, 3) real-time, ECG assisted detection, 4) real-time, ECG independent detection.

In the best conditions (offline, ECG assisted detection), 97% of maximum ventricular volume and 95.7% of minimum ventricular volume were accurately detected over 27 seconds of TCG signal monitoring.
In the worst conditions (real-time, ECG independent detection), 87.2% of maximum ventricular volume were detected, against 92.7% of minimal ventricular volume.

Despite a reduced efficiency, the algorithms did not produce errors, i.e. no event in the TCG signal was mistaken for an extremum ventricular volume.

The detection algorithms also permitted to precisely study of the position in time of the extrema ventricular volume, in relation with the R wave of the ECG signal. As expected, it resulted that these events observe a consistent position in time[6]. This particularity is used to facilitate the detection of maximum and minimum ventricular volume.

---

[6] Result in accordance with physiology prove that the FORP is capable of accurate measurement of cardiac motion.

Chapter 13 CONCLUSIONS AND FURTHER WORK

13.1. Overview

The aim of this investigation was to enhance the characteristics and capabilities of the FORP. The research presented in this thesis has had positive outcomes from both instrumental and medical points of view, and was focused on four specific areas of fibre optic plethysmography. These areas are the sensor itself, the acquisition system, the application of plethysmography to respiratory monitoring and to cardiac monitoring.

These last two areas are the main issues in plethysmography at the moment. It is believed that they have been successfully addressed by providing original solutions. High resolution cardiac and respiratory monitoring is now achievable with the new FORP prototype, thanks to its improved specifications.

It is hoped that the present research will help in establishing fibre optic plethysmography as a sound medical investigation technique. The findings in each area of the research are summarised below:

13.1.1. Equivalence of respiratory volume estimation with the FORP and the RIP

It was demonstrated in section 7.6.3 that, despite the conceptual differences of the FORP and the RIP, these two plethysmographic systems can produce equivalent estimation of the respiratory volume signal (as measured with a spirometer or a pneumotachometer) during quiet breathing, i.e. when the magnitude of thoracic and abdominal excursions is limited. This is true *a fortiori* during cardiac position monitoring since the TCG signal peak-to-peak amplitude is smaller than 5% of the peak-to-peak respiratory amplitude.

13.1.2. Redevelopment of the sensor

A new sensor, based on a simple coil configuration resembling a figure-of-eight knot, has been developed with success, and was presented in chapter 9. For the same elongation, the response of the new sensor is 2 to 3 times greater than the previous FORP sensor used in A. Raza's research, while showing a more linear behaviour.

The figure-of-eight configuration helps to reduce the sensor hysteresis.

A simple geometrical model of the figure-of-eight has been developed and compared against experimental results. An estimation of the figure-of-eight behaviour based on this model has good precision and reasonable accuracy.

The sensor's experimental optical response to elongation has been compared to an existing model of the MBLE with good agreement, except for small curvature radii.
Despite this discrepancy, both model and experimental results can be safely approximated as linear responses during small elongation variation of the figure-of-eight, as would be the case during quiet breathing, and a fortiori during cardiac monitoring while holding breath.

13.1.3. Redevelopment of the acquisition system

The new acquisition system presented in chapter 10 was designed around the idea of synchronous emission and acquisition of the optical signal travelling through the two plethysmographic sensors. Subsequently, the light source could be pulsed at any desired frequency or duty factor without loss of synchronisation.

The system is PC based, and permits the acquisition of other signals such as the ECG, the signal from a spirometer or the signal from a microphone if one's intention is to record heart sounds.

A Tailor made acquisition system was developed for the PC and the control application was written in C for DOS.

13.1.4. New calibration method for respiratory monitoring

The problem of accurate and precise calibration of the FORP against a spirometer was addressed with a new approach in chapter 11.

A calibration method, based on independent filtering of the thoracic and abdominal signals with a Wiener filter linear combiner, has proven successful by dramatically reducing the error between the spirometric signal and its plethysmographic estimation by 79%, compared to the previously most efficient calibration method (regressive calibration of quadratic respiratory model).

In this approach, the spirometer signal is not modelled, but simulated since the filtering operations have no physiological grounding.

13.1.5. Cardiac movements monitoring for medical scanners

Thanks to the improved sensor and acquisition system, the generation of high quality plethysmographic signals has revealed the capability of the FORP to pick up body circumference variations due to cardiac activity on different body parts such as the cranium, neck, thorax and abdomen. This is reported in chapter 12.

The plethysmographic cardiac activity observed at thoracic level interested us in particular because, according to previous investigations of plethysmographic cardiac monitoring, it is directly related to the heart movements if the sensor is correctly placed on the chest. For this reason, the FORP offers a potentially better cardiac movement tracking capability than the ECG. Better synchronisation of the heart to the scanner should permit an increase in the image quality recorded by the scanner.

It was demonstrated that the maximum and minimum ventricular volume of the cardiac cycle could be successfully detected if the patient is able to provide a stable chest position in apnoea.

During an acquisition session of 26 seconds, the detection of maximum ventricular volume was 97% successful (3% missed) with offline processing and a parallel ECG signal to assist detection. Without ECG assistance, the success rate was 86.1% (13.9% missed).

We also demonstrated that real time monitoring of cardiac movements with the FORP was possible. This capability was developed in view of monitoring the heart's position during cardiac imaging with MR and/or CT scanners.

In real time, the detection of maximum ventricular volume was 87.2% successful (12.8% missed) when using a parallel ECG signal to assist detection. Without ECG assistance, the success rate of real time detection was 75.5% (24.5% missed). Although this last figure is relatively high, it must be noted that the algorithm never produced errors, i.e. heartbeats were missed but not wrongly detected. As a result, it would be safe to use this algorithm, even if it is not 100% efficient at detecting maximum ventricular volume.

13.2. Further work

It is the nature of research to be an everlasting quest, often asking more questions than actually solving problems. The numerous developments in the FORP system presented in this thesis have raised new issues of technical and scientific relevance concerning fibre optic plethysmography. Further work is required to carry on improving the quality of plethysmographic signals and proving the applicability of fibre optic plethysmography to cardiac and respiratory monitoring in clinical settings.

13.2.1. Improving the fibre optic plethysmographic system

Numerous immediate technical improvements of the FORP are envisaged in order to produce yet more precise and accurate results, which is highly desirable for cardiac monitoring. In particular, to reduce the noise level of plethysmographic signals at their very source (the optronic box, see section 10.3), different solutions are immediately available:

- The two IR-LED driver circuits should be dissociated from the photocurrent amplifier circuits. A present, these two circuits share the same PCB, which is unfortunate because the IR-LED drivers generate large pulsed currents that can disturb the nanoamps-level photodiode currents.
- This spatial separation would also allows to experiment with more powerful optical sources (such as laser diodes) without impairing the quality of plethysmographic signal. New optical sources could be powered with rechargeable batteries so as to remove mains coupling. Nickel-Cadmium batteries for example are perfectly capable of delivering repeated pulsed currents of a few Amps.
- The photodiode amplifier circuits could also be powered using batteries and the output signals could be opto-isolated from the acquisition card so as to prevent propagation of power line pollution due to the computer.
- Finally, the quality of the optical fibre connections to the IR-LEDs and photodiodes could be greatly improved by using professional connectors, instead of plastic screw-in plugs.

To operate in real-time, the acquisition system and the signal processing algorithms must be merged within the same computer. At present, the acquisition computer suffers from a serious lack of computational power (the processor is a "prehistoric" Intel 486DX4) and therefore is not capable of performing signal processing in a reasonably short amount of time. A fortiori, real time DSP is out of the reach of this machine.

A faster computer should be used and, if cardiac position monitoring with the FORP is to be tested in different hospitals, it would be desirable to facilitate its transport and installation.

Ideally, the FORP system's hardware should be reduced to a single module that could be interfaced with a parallel port[1] to any Windows PC (say) thanks to a single application regrouping control operations and signal processing.

Alternatively, the solution of using a specific DSP chip should be considered if further developments see the need for a smaller system.

13.2.2. Further testing of the optimal filters calibration method for respiratory volume estimation

It was established that calibration by optimal filtering of the plethysmographic signals permitted better estimation of the respiratory volume signal. This should be further confirmed by applying the calibration method to a larger number of subjects, in order to increase the statistical significance of results. In order to do so, new experimental protocols should be designed to enable the simultaneous uses of the FORP and a spirometer on different population groups (young/old, male/female, etc.) in different conditions (supine/seating/standing, healthy/unhealthy, sleeping/awake, etc.).
The frequency of recalibration needs to be determined, and, more importantly perhaps, what are the factors that makes it necessary.

13.2.3. Improvements to the signal processing aspect of the optimal filters calibration method

A few immediate modifications were already suggested in section 11.5. so as to speed up the calibration process and convert the system to real time operation. This essentially requires decreasing the sampling frequency and increasing the computational power of the system.

Equally important was the exciting possibility that, in order to reduce the computational cost of determining coefficients values for the optimal filter, an LMS adaptive filter could be used. Preliminary tests have been inconclusive but it is believed that this technique has fairly good chances of success.

13.2.4. Improvement to cardiac detection algorithm

The main drawback of cardiac position monitoring with the FORP is the limited ability of the signal processing algorithm to resolve maximum ventricular volume in real-time without assistance from a parallel ECG signal.

---

[1] Both hardware and software aspect of the communication protocol in use with the 8 bits parallel port are extremely stable on the PC platform. An external system communicating with this port is virtually 100% compatible to any PC. The parallel port can be simply accessed within any operating system.

A simple solution to this problem would be to replace the ECG signal by the plethysmographic signal obtained around the neck of the patient. It was shown in chapter 12, section 12.3, figure 2 that such a signal has a clear, distinctive shape with a consistent and easily recognisable maximum corresponding to the maximum blood pressure during the cardiovascular pulse following ventricular ejection. Moreover, this plethysmographic signal has a much smaller dependence on spurious respiratory movements and therefore should be simpler to process algorithmically.
This method should hopefully allow us to equate the quality of ECG assisted detection (higher efficiency), but without the ECG signal.

13.2.5. Testing the hypothesis that medical scanners synchronised to TCG signals would permit better cardiac imaging

In order to verify the assumption presented in this thesis that the TCG signal could help to produce better cardiac imaging by detecting intervals of minimal heart movements, different experiments should be designed where a scanner could be trigged either by TCG or ECG synchronisation. Such experiment could not be carried by the author because experimental use of MR and CT scanners is not only extremely time consuming, but also requires the presence of a subject, a scanner operator and a specialised clinical practitioner at any time. The cost of such experimentation is probably very high, given its dependence on medical staff and medical equipment.

However, to prove the hypothesis of TCG synchronised cardiac imaging, the scanner might simply be replaced by an echocardiographic system. Using this system, one might be able to verify that the heart movements are correlated with the TCG signal and, more importantly perhaps, that the detection of regions of minimal cardiac movements using the stability estimator presented in section 12.4.1.4, is an accurate technique.

REFERENCES

Agostini E., Mognoni P., Torri G., Saracino F. 1965a Relation between changes of rib cage circumference and lung volume. J. App. Physiol. 20: pp.1181-1187.

Agostini E., Mognoni P., Torri G., Saracino F. 1965b Static features of the ribcage and abdomen diaphragm. J. App. Physiol. 20: pp.1181-1187.

Allison R. D., Holmes E. L., Nyboer J. 1964 Volumetric dynamics of respiration as measured electrical impedance plethysmography. J. App. Physiol. 19: pp.166-176.

Almond N. E., Cooke E. D. 1989 Observation on PPG pulse derived from laser doppler flow meter. Clin. Phys. Physiol. Meas. 10: pp.137-145.

Ansari F. and Navalurkar R. K. 1993 A Fiber optic sensor for the determination of dynamic fracture parameters in fibre reinforced concrete in Application of Fiber optic sensors in engineering mechanics. Proc. Am. Soc. Civ. Eng. pp. 160-176.

Augousti A. T. 1997 A theoritical study of the robustness of the isovolume calibration method for two compartment model of breathing, based on an analysis of the connected cylinders model. Phys. Med. Biol. 42: pp. 283-291.

Augousti A. T. and Raza A. 1993 The development of a fibre optic respiratory plethysmograph. Sensors VI: Technology, Systems and Applications proc. pp.401-406.

Augousti A. T., Raza A., Graves M.1996 Design and characterisation of a fibre optic respiratory plethysmograph. BiOS 96 - Biomedical Sensing, Imaging and Tracking Technologies proc. pp.250-257.

Augousti A.T., J.Mason, Maletras F.-X. 2001 The Fibre Optic Respiratory Plethysmograph - an overview. (invited presentation) Institute of Physics Annual Congress, Brighton 19.3.2001.

Babchenko A., Khanoch B., Shomer Y., Nitzan M. 1999a Fibre optic sensor for the measurement of respiratory chest circumference changes. J. Biomed. Opt. 4: pp. 224-229.

Babchenko A.,Turinvenko S., Khanoch B., Nitzan M. 1999b A fibre optic sensor for the measurement of breathing effort. SPIE proc. 2631. pp.64-71.

Bazett H. C. 1920 An analysis of the time relation of electrocardiograms. Heart 7: pp.353-370.

Bellanger M. 1994 Digital Processing of Signals (second edition). Wiley Interscience.

Bendat J. S. and Piersol A.G. 1986 Random Data, Analysis and Measurement Procedures (second edition). Wiley-Interscience.

Bloch K. E., Juggon S., Sackner M. A. 1994 Thoracocardiographic derived left ventricular systolic time intervals. Chest 106: pp. 1668-1674.

Bloch K. E., Jugoon S., De Socarraz H., Manning K. E., Sackner M. A.1998 Thoracorcardiography: Non-invasive monitoring of left ventricular stroke volume. J. Crit. Care 13: pp.146-157.

Bloch K. E., Jugoon S., Sackner M. A. 1999 Thoracocardiography: Non-invasive monitoring of left ventricular filling. J. Crit Care 14: pp. 177-185.

Boechat A.A. P., Su D., Hall D.R., Jones J.D.C. 1991 Bend Loss in large core multimode optical fiber beam delivery system. Applied Optics, 30: pp.321-327.

Briers J. D. 1996 Waves and Optics. Kingston University, Kingston-upon-Thames, U.K.

Briers J. D. 1997 Optical theories and their applications. Kingston University, Kingston-upon-Thames, U.K.

Brown B. H. Some new instruments for the continuous monitoring of body temperature, respiratory rate and pulse rate. 1966 Phys. Med Biol. 11: pp.135-137.

Brown K., Aun C., Jackson E., Mackersie A., Hatch D., Stocks J. 1998 Validation of RIP using the QDC method in anaesthetized infants. Eur. Respir. J.

Burr-Brown 1995 Photodiode monitoring with Op Amps, application note AB-075. Burr-Brown.

Chada T., Watson H., Birch S., Jenouri G., Scneider A., Cohn M., Sackner M. 1982 Validation of RIP using different calibration procedures. Am. Rev. Resp . Dis.

Challoner A. V. J. 1979 Photoelectric plethysmography for estimating cutaneous blood flow, in Non-invasive physiological measurement, editied by P. E. ROLFE., Academic Press, London. 125-151.

Clancy J., McVicar A. J. 1995 Physiology and Anatomy (seconde edition). Arnold.

Cohn M. A., Rao A., Broudy M., Birsh S., Watson H., Atkins N., Davis B., Stott F., Sackner M. 1982 The Respiratory Inductive Plethysmograph: A new non-invasive monitor of respiration. Bull. Europ. Physiopath. Resp. 18: pp.643-658.

Cohn M. A., Watson H., Weisshaut R., Stott F., Sackner M. A. 1978 A transducer for non-invasive monitoring of respiration. Proc. of the second international symposium on ambulatory monitoring. pp.119-128.

Costanzo L. S. 2002 Physiology (second edition). Saunders.

Crawford A. B. H., Dodd D., Engel L. A. 1983 Change in ribcage shape during quiet breathing, hyperventilation and single inspiration. Resp. Physiol. 54: pp.197-209.

Cronin N. J. 1995 Microwave and optical waveguide. IOP Publishing.

Crooks L. E., Arakwa M., Hoenniger J. 1982 NMR whole body imager operating at 3.5 Kgauss. Radiology 143: pp.169.

Davis C., Mazzolini A., Mills J., Dargaville P. 2000 A new sensor for monitoring chest wall motion during high frequency oscillatory ventilation. Med. Eng. & Phy. 21: pp.619-623.

Davis C., Mazzolini A., Murphy D. 1997 A new fibre optic sensor for respiratory monitoring. Autralasian Physical & Engineering Sceinces in Medecine. 20: pp.214-219.

Dorlas J. C., Nijboer J. A. 1985 Photoelectric plethysmography as a monitiring device in anaesthesia. Ibid. 57: pp.524-530.

Fleisch A.1925 Der Pneumotachograph: ein apparat zur geshwindigkeitdregistrirung der Atemluft. Arch. Ges. Physiol. 209: pp.713-722.

Fox S. I. 1999 Human Physiology (sixth edition). MacGrow-Hill.

Ghatak A., Thyagarajan K., 1998 Introduction to Fiber Optics. Cambridge Uni Press.

Gloge D. 1972 Bending Loss in Multimode Fibers with Graded and Ungraded Core Index. Applied Optics. 11: pp. 2506-2513.

Gloge D.1971 Weakly Guiding Fibers. Applied Optics. 10: pp.2252-2258.

Gosbee J., DeRosier J. 2001 MR Hazard Summary. www.patientsafety.com

Grant P., Cowan C., Mulgrew B., Dripps J. 1989 Analogue and Digital Signal Processing and Coding. Chartwell-Bratt Studentlitteratur.

Graves M., Raza A., Augousti A.T. 1996 A fibre optic respiratory monitoring and feedback system for MRI. Proc. Radiology 96.

Guyton A. C., Hall J. E. 2001 Textbook of medical physiology (tenth edition).

Haldane J. S. 1905 The regulation of lung ventilation. J. Physiol. 32: pp.242.

Hamilton L. H., Beard J. D., Carmean R. E., Kory R. C. 1967 An electrical impedance ventilometer to quantitate tidal volume and respiration. Med. Res. Eng. 6: pp.11-16.

Haykin S. 1991 Adaptive Filter Theory (second edition). Prentice Hall.

Hecht E. 1998 Optics (third edition). Addison-Wesley.

Henneberg S., Hok B., Wiklund L., Sjodin G. 1992 Remote ausculatory patient monitoring during MRI. J. Clin. Monit. 8: pp.37-43.

Hornack Joseph P. 2000 The basics of MRI.

Horowitz P. and Hill W. 1980 The Art of Electronics. Cambridge University Press.

Ifeachor E. and Jervis B. 1993 Digital Signal Processing, a practical approach. Addison-Wesley.

Jensen N. F. 2001 Big Red, anesthesiology oral board preparation. www.boardprep.com

Jones K. A.1987Introduction to optical electronics. John Wiley.

Jordan C., Henke K., Stone A., Brandwayn L., Belsito A., Sackner M. 1984 Measurement of systolic time intervals during exercise using inductive plethysmography. Int. J. Clin. Monit. Comput. 1: pp.137-146.

Kamal A. A. R., Harness J. B., Irvin G., Mearns A. J. 1989 Skin photoplethysmography - a review. Comp. Meth.& Prog. in Biomed. 28: pp.257-259.

Kauppinen S., Hyttinen J., Koobi T., Kaukinen S., Malmivuo J. 2001 Impedance Cardiography. Int. J Bioelectromagnetism.
http://ee.tu.fi/rgi/ijbem/volume3/number2/kaupinen

Kelly R., Hayward C., Ganis J., Daley J., Avolio A., O'Rourke M. 1989 Non-invasive registration of the arterial pressure pulse waveform using high-fidelity applanation tonometry. J. Vasc. Med. Biol. 1: pp.142-149.

Konno K., Mead J. 1967 Measurement of the separate volume changes of rib cage and abdomen during breathing. J. App. Physiol. 22: pp.407-422.

Koritke J. G., Sick H. 1988 Atlas of sectional human anatomy (second edition). Urban & Scharzenberg.

Kubicek W.G., Karnegis J. N., Patterson R. P., Witsoe D. A., Mattson R. H.1966 Development and evaluation of an impedance cardiac output system. Aerospace Medecine. 37: pp.1208-1212.

Lee J. 2000 ECG monitoring in theatre. Update in anaesthesia. Issue. 11 Art. 5.

Lindberg L. G., Ugnell H., Oberg P. A. 1992 Monitoring of respiratory and heart rate using a fibre-optic sensor. Medical & Biological Engineering & Computing 30: pp.533-537.

Lynn P. 1982 An introduction to the analysis and processing of signals. The MacMillan press ltd.

Maletras F.-X., Augousti A.T., Mason J. 2001a Construction and calibration of a new design of fiber optic respiratory plethysmograph (FORP). SPIE 4444 Proc. of Optomechanical Design and Engineering. pp285-293.

Maletras F-X, Augousti A.T., Mason J. 2000 Combined cardiac and respiratory monitoring using the Fibre Optic Respiratory Plethysmograph (FORP). Optics and Optoelectronics 2000, Loughborough.

Maletras F-X, Augousti A.T., Mason J. 2001b Signal processing considerations in the use of the fibre optic respiratory plethysmograph (FORP) for cardiac monitoring. Proc. of Sensors and Their Applications XI. pp.371-376.

Martinot Lagarde P., Sartene R., Mathieu M., Durand G. 1988 What does inductance plethysmography really measure ? J. App. Physiol. 64: pp.1749-1756.

McCool F. D., Kelly K. B., Loring S. H., Geaves I. A., Mead J. 1986 Estimates of ventialtion from the body surface measurements in unrestrained subjects.

Mead J., Peterson N., Grimby G. 1967 Pulmonary ventilation measured from body surface movements. Science 9: pp.1383-1384.

Newman D. G., Callister R. 1999 The non-invasive assessment of stroke volume and cardiac output by impedance cardiography: A review. Av. Spa. Env. Med. 70: pp. 780-789.

Noordergraaf A., Burger H. C., Starr I. 1956 Physical basis of ballistocardiography. PhD thesis. University. Of Utrecht, Netherlands.

Nyboer J. 1970 Electrical Impedance Plethysmography (second edition). C C Thomas, Springfield.

Patterson R. P., Kubicek W. G., Kinnen E., Witsoe D. A., Noren G. 1964 Development of an electrical impedance plethysography system to monitor cardiac output. Proc. Of the 1st annual rocky mountain bioengineering symposium. pp.56-71.

Pocock G., Richards C. D. 1999 Human Physiology, the basis of medecine. Oxford University Press.

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992 Numerical Recipes in C, the art of scientific computing (second edition). Press syndicate of the University of Cambridge.

Raza A. 1998 Development and Construction of a Fibre Optic Respiratory Plethysmograph (FORP). PhD thesis, Kingston University, Kingston-upon-Thames.

Raza A., Augousti A. T., Graves M. 1995 Optical instrumentation for respiration measurement in magnetic resonance scanners. 7th IEE Proc. Progress in Fibre Optic Sensors and Their Applications. pp.15/1-15/5.

Raza. A., Augousti. A.T. 1994 Further developemt of a fibre optic respiratory plethysmograph. Proc. Of App. Opt. & OptoElec. Conf. pp. 271-273.

Reynaud R. 1995 Cours de Traitement du Signal. Universite Paris-Sud.

Roth C. 1996 MR Safety. Outsource, Inc.

Sackner M. A., Hoffman R. A., Stroh D. 1991 Thoracocardiography part I: Non-invasive measurement in stroke volume. Chest 99: pp.613-622.

Sackner M. A., Watson H., Belsito A., Feinerman D., Suarez M., Gonzalez G., Bzouski F., Krieger B. 1989 Calibration of respiratory inductive plethysmogrpah during naturla breathing. J. App. Physiol. 66: pp. 410-420.

Sanchez-Ortiz G. I., Burger P. 1996 A Novel Gating Technique for Cardiac NMR Imaging using Heartbeat Time Intervals Forecast.

Schild H. H. 1993 MRI made easy.

Senior J. 1984 Optical Fiber Communications. Prentice Hall.

Sharp J. T., Goldberg N. B., Druz W. S., Danon J. 1975 Relative contributions of rib cage and abdomen to breathing in normal subjects. J. App. Physiol. 39: pp.608-618.

Smith J., Mead J. 1981 Degrees of freedom in chest wall motion in man. Physiologist 24: pp.539.

Sullivan W. T., Peters G.M., Enright P.L. 1984 Pneumotachograph: Theory and clinical applications. Respirat. Care 29: pp.736-749

Vander A. J., Sherman J. H., Luciano D. S. 1994 Human Physiology, the mechanisms of body functions (sixth edition). McGrow-Hill.

Wade D. L. 1954 Movement of the thoracic cage and diaphragm in respiration. J. Physiol. 124: pp.193-212.

Welch, P.D. 1967 "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms". IEEE Trans. Audio Electroacoustics. Vol. AU-15 (June 1967). pp. 70-73.

Werthamer J., Krasner J., Dibenedetto J., Startk A. R. 1983 Apnea monitoring by acoustic detection of air flow. Pediatrics 71: pp.53-55.

Widrow B., Glover J. R., McCool J. M., Kaunitz J., Williams C.S., Hearn R.H., Zeidler J.R., Dong E., Goodin R.C. 1975 Adaptive noise cancelling: Principles and applications. Proc. IEEE, 63, 1692-1676

Wilhelm F. H., Roth W. T. 1996 Ambulatory assessment: computer-assisted psychological and psychophysiological methods in monitoring and filed studies. Ambulatory assessment of clinical anxiety, by J. Fahrenberg and M. Myrtek. Gottingen: Hogrefe & Huber.

Wilhelm F. H., Roth W. T., Sackner M. A. 2001 The LifeShirt: An advanced system for ambulatory measurement of respiratory and cardiac functions. Pres. at the International society for the advancement of respiratory psychophysiology, 14th-16th september 2001, Oxford.

Zimmerman P. V., Connellan S. J., Middleton H. C., Tabona M. V., Golman M. D., Pride N. 1982 Postural changes in rib cage and abdominal volume motion coefficients and their effect on the calibration of a RIP. Am. Rev. Resp. Div. 127: pp.209-214.

APPENDIX 1 LISTING OF THE FORP SYSTEM CONTROL APPLICATION

```
//-----------------------------------------------------------------
//                        SYNACQ23
//-----------------------------------------------------------------

/*

This programme produces acquisition of 4 signals on the prototype
acquisition card.


Pin map of the IO/timer card

      clock       on board clock Fc = 1MHz

      chip 8254

      counter0    SQW prescaler, generate Fs = 2KHz
      counter1    diode pulse, trigged by counter0 output
      counter2    sh pulse, trigged by counter0 output

      chip 8255 #1 in mode 1 with PORTA input PORTB output
      PORTB       B7    output flag ISR
      PORTC       C3    input from 8253 counter1
                  C4    outpur to bus IRQ

      chip 8255 #2 in mode 0 with PORTA output PORTB imput PORT input

      PORTA       A0    EN     DG409       Enable
                  A1    A0     DG409       Address Line 0
                  A2    A1     DG409       Address Line 1
                  A3    CS     MAX172      Chip Select
                  A4    RD     MAX172      Read
                  A5    HBEN   MAX172      Enable D8-D11 on D0-D3
      PORTB       B0    D0/8   MAX172      Data Lines
                  B1    D1/9   MAX172
                  B2    D2/10  MAX172
                  B3    D3/11  MAX172
                  B4    D4     MAX172
                  B5    D5     MAX172
                  B6    D6     MAX172
                  B7    D7     MAX172
      PORTC       C0    BUSY   MAX172      Status

*/
```

```
/* LIBRARIES ------------------------------------------------ */

#include <conio.h>
#include <stdlib.h>
#include <stdio.h>
#include <dos.h>
#include <alloc.h>
#include "synacq11.h"   // All constants relating to SYNACQ

/* TYPE DEFINITIONS ----------------------------------------- */

typedef unsigned long int DWORD;
typedef unsigned char BYTE;
typedef unsigned int WORD;

/* GLOBAL VARIABLES ----------------------------------------- */

float Fs=2000,pulse=8.6,sh=7.6;
DWORD duration,numberofsamples;
WORD huge *datavect;
int memorymap[4], channellist[4], numberofchannels=0;
char logfilename[80] = "d:\\data\\logfile.txt";
char **filename;


/* INTERRUPTION RELATED "PSEUDO-GLOBAL" VARIABLES ----------------*/

BYTE dummy,i,j=0,m,mpxaddressmask;
DWORD index;
WORD lowerbyte, upperbyte, measure;

/* FUNCTIONS PROTOTYPES ------------------------------------------*/

void requestarguments(void);        // if no inline argumemts when call
to main then request arguments from user
void displayarguments(void);        //
void showmemory(void);              // analyse available conventional
memory
int  allocatememory(DWORD samples); // dynamic memory allocation void
boardsinit(void);                   // initialise all hardware resources
void reset8253(void);               // put counters outputs high
void acquisition(DWORD samples);    // set IRQ/do nothing/reset IRQ after
acq.
void savedata(void);                // prepare for saving data files
int  data2file(DWORD samples);      // transfer datavect to text file
void wait(void);                    // WAIT !!!!
void beep(void);                    // BEEP !!
void interrupt acquire(void);       // present hardware ISR
void interrupt (*old)(void);        // old hardware ISR
```

```
/* MAIN FUNCTION  ------------------------------------------------*/

int main(int argc, char *argv[])
{
char anykey;
clrscr();
printf("\n");
numberofchannels = 0;
for (i=0;i<4;i++)
      {
      channellist[i] = NO;          // initialise lists
      memorymap[i] = -1;
      }
if(argc<2) requestarguments();
else  {
      duration = atoi(argv[1]);     // sampling duration
      for (i=0;i<4;i++)
            {
            if (atoi(argv[i+2]) > 0)
                  {
                  channellist[atoi(argv[i+2])-1] = OK;
                  memorymap[atoi(argv[i+2])-1] = numberofchannels;
                  numberofchannels += 1;
                  }
            }
      }
reset8253();
boardsinit();                       // initialise 8253 and 8255
```

```
for(;;)
{
clrscr();
displayarguments();
numberofsamples = Fs * duration;          // compute number of samples
if(allocatememory(numberofsamples))       // if memory allocation
succesfull.
      {
      printf("\nPress any key to start acquisition...");
      getch();
      printf("\r                                        ");

      printf("\rAcquisition...\t\t");
      beep();

      acquisition(numberofsamples);
      printf("done");

      beep();
      savedata();

      free(filename);
      farfree(datavect);                  // memory is released after use
      } // end of if(allocatememory)

printf("\nRun again ?  (y/n)\t");
anykey = getch();
anykey = toupper(anykey);
if(anykey=='N')
      {
      printf("No");
      break;
      }
else  {
      printf("Yes\n");
      numberofchannels = 0;
      for (i=0;i<4;i++)
            {
            channellist[i] = NO;          // initialise lists
            memorymap[i] = -1;
            }
      requestarguments();
      }
} // end of infinite for loop

printf("\nReset 8253 ? (y/n)\t");
anykey = getch();
anykey = toupper(anykey);
if(anykey=='Y')
      {
      reset8253();                        // make sure 8253 OUTs are high
      printf("Yes");
      }
else printf("No");


printf("\nEnd");
if(argc<2) getch();
return OK;
}
```

```
/* OTHER FUNCTIONS DEFINITIONS ----------------------------------------*/

void requestarguments(void)
{
char anykey,confirm;
numberofchannels = 0;
for (i=0;i<4;i++)
      {
      channellist[i] = NO;          // initialise lists
      memorymap[i] = -1;
      }
printf("Sampling duration (s) :\t");scanf("%lu",&duration);
for(i=0;i<4;i++)
      {
      printf("Channel %i ? (y/n)\t",i+1);
      anykey = getch();
      anykey = toupper(anykey);
      if (anykey=='Y')
            {
            printf("Yes\n");
            channellist[i] = OK;
            memorymap[i] = numberofchannels;
            numberofchannels +=1;
            }
      else  {
            printf("No\n");
            }
      }
} /* end of requestarguments */


void displayarguments(void)
{
char answer[4];
printf("\nSampling Freqency...\t%4.0f Hz",Fs);
printf("\nSampling Duration...\t%i s", duration);
printf("\nDiode Pulse...\t\t%2.1f æs",pulse);
printf("\nSample & Hold Pulse...\t%2.1f æs",sh);
for(i=0;i<4;i++)
      {
      if (channellist[i] == OK)
            {
            strcpy(answer,"Yes");
            }
      else  {
            strcpy(answer,"No");
            }
      printf("\nChannel %i...\t\t%s",i+1,answer);
      }

} /* end of displayarguments */


void showmemory(void)
{
DWORD memory;
memory = farcoreleft()/1024;  // determine system free memory for
allocation
printf("\nSystem resources...\t%luk",memory);
} // end of showmemory
```

```
int allocatememory(DWORD samples)
{
DWORD memory;
memory = farcoreleft()/1024;  // determine system free memory for
allocation
if((memory>640)||(memory<0))
      {
      printf("\nMemory error");
      return NO;
      }
printf("\nSystem resources...\t%luk",memory);
printf("\nMemory
request...\t%ik",(numberofchannels*samples*sizeof(WORD))/1024);
printf("\nMemory allocation...\t");
// determine base address of data array
datavect  = (WORD*)farmalloc(numberofchannels*samples*sizeof(WORD));
if(!datavect)
      {
      printf("Not enough memory");
      return NO;
      }
printf("done");
printf("\nMemory initialise...\t");
for(index=0;index<(numberofchannels*samples);index++) datavect[index] =
0;
printf("done");
return OK;
} // end of allocatememory()


void boardsinit(void)
{
DWORD Ns=0,NsLSB=0,NsMSB=0,Np=0,NpLSB=0,NpMSB=0,Nsh=0,NshLSB=0,NshMSB=0;

pulse = pulse + 1.4;                  // correction on diode pulse duration
sh = sh + 1.4;                        // correction on sh pulse duration

// Time Constants Determination
Ns = (int)(Fc / Fs);                  // counter0 divider
NsMSB = (int)(Ns / 256);
NsLSB = Ns - (256*NsMSB);
Np = (int)((Fc*pulse)/1000000);       // counter1 divider
NpMSB = (int)(Np / 256);
NpLSB = Np - (256*NpMSB);
Nsh = (int)((Fc*sh)/1000000);         // counter2 divider
NshMSB = (int)(Nsh / 256);
NshLSB = Nsh - (256*NshMSB);

// initialising 8253 Counters
outportb(COUNTERCONTROL,SQW0);        // initialise le 8253 counter0 en
mode square wave genrator
outportb(COUNTER0,NsLSB);             // load divider lsb for sampling
frequency
outportb(COUNTER0,NsMSB);             // same but msb
outportb(COUNTERCONTROL,SHOT1);       // initialise le 8253 counter1 en
                                      mode one // shot (diode pulse)
outportb(COUNTER1,NpLSB);             // load divider lsb for pulse period
outportb(COUNTER1,NpMSB);             // same but msb
outportb(COUNTERCONTROL,SHOT2);       // initialise le 8253 counter2 en
                                      mode one // shot (sample&hold)
outportb(COUNTER2,NshLSB);            // load divider lsb for pulse period
outportb(COUNTER2,NshMSB);            // same but msb

// Initialising 8255 IO chips
outportb(IOCONTROLPORT,MODE1AINBOUT);      // initialise le 8255 #1
outportb(IOCONTROLPORT2,MODE0AOUTBINCIN); // initialise le 8255 #2
```

```
// Initialising MAX172 AD chip
outportb(IOPORTA2,inportb(IOPORTA2)&(~CS));// select chip
outportb(IOPORTA2,inportb(IOPORTA2)|HBEN );     // set HBEN
outportb(IOPORTA2,inportb(IOPORTA2)|RD);        // set RD
outportb(IOPORTA2,inportb(IOPORTA2)|CS);        // deselect chip

// Initialising DG409 analog multiplexer to AI0
outportb(IOPORTA2, inportb(IOPORTA2)|EN);       // select DG 409
outportb(IOPORTA2, inportb(IOPORTA2)&(~A0));    // reset A0 select line
outportb(IOPORTA2, inportb(IOPORTA2)&(~A1));    // reset A1 select line
//outportb(IOPORTA2, inportb(IOPORTA2)&(~EN));  // deselect DG409

} // end of boardsinit
```

void reset8253(void)

```
// make sure the OUTs are high to protect the IRdiode from continuous
//power feeding that would blow it up
{
outportb(COUNTERCONTROL,RESET2);    // stop counter2 by
// counting untill 0x01 so that OUT2 is high to switch off IR diode (so
that it //doesn't burn...)
outportb(COUNTER2,1);
outportb(COUNTER2,0);
outportb(COUNTERCONTROL,RESET1);    // stop counter1 by
// counting untill 0x01 so that OUT1 is high to put S&H amp in sample
mode (so //that the S&H output is static)
outportb(COUNTER1,1);
outportb(COUNTER1,0);
outportb(COUNTERCONTROL,RESET0);    // stop counter0, OUT0 can stay low
} // end of reset8253
```

void acquisition(DWORD samples)

```
{
index = 0;                          // initialise interruption counter
old = getvect(8+INTNUM);            // get old ISR from IRQ
setvect(8+INTNUM, acquire);         // assign IRQ to new ISR acquire
enable();                           // enable all IRQs
outportb(IOCONTROLPORT,INTCE);      // enable IOPORTC IRQ on strobe
outportb(IMR,inportb(IMR)&(~IRQ));  // demask IRQ
for(;;)     {
      if(index>=samples) break;     // do nothing untill acquisition
finished
      } /* end of for */
outportb(IOCONTROLPORT,INTCD);      // disable IOPORTC IRQ on strobe
outportb(IMR,((inportb(IMR))|IRQ)); // remask IRQ
setvect(8+INTNUM,old);              // replace old interruption routine
} // end of process
```

```
void savedata(void)
{
char anykey;

filename = (char*)malloc(4*sizeof(char)); // assign pointer
for(i=0;i<4;i++)
      {
      filename[i] = (char*)malloc(80*sizeof(char));
      }
filename[0] = "d:\\data\\data0.txt";
filename[1] = "d:\\data\\data1.txt";
filename[2] = "d:\\data\\data2.txt";
filename[3] = "d:\\data\\data3.txt";

printf("\nSaving files ? (y/n)\t");
anykey = getch();
anykey = toupper(anykey);
if(anykey=='Y')
      {
      printf("Yes");
      printf("\nUse default names ?(y/n)");
      anykey = getch();
      anykey = toupper(anykey);
      if(anykey=='Y') printf("Yes\n");
      else    {
            printf("No\n");
            for(i=0;i<4;i++)
                  {
                  if(channellist[i] == OK)
                        {
                        printf("Path and filename %i ?\t",i+1);
                        scanf("%s", filename[i]);
                        }
                  }
            }
      printf("Saving files...\t\t");
      if(data2file(numberofsamples)) printf("done");
      }
else  {
      printf("No");
      }
} //end of savedata

int data2file(DWORD samples)
{
FILE *data[4], *logfile;
struct date d;
struct time t;

getdate(&d);
gettime(&t);


logfile = fopen(logfilename,"wt");
if(!logfile)
      {
      printf("\nError opening logfile");
      return NO;
      }

for (i=0;i<4;i++)
      {
      if(channellist[i] == OK )
            {
            data[i] = fopen(filename[i],"wt");  // mode is write as text
            if(!data[i])
```

```
                    {
                    printf("\nError opening file %i",i);
                    return NO;
                    }
              else  {
                    // OK for saving file
                    fprintf(logfile,"%i\n",i+1);
                    for(index=0;index<samples;index++)
                          {
                          fprintf(data[i],"%i\n",datavect[index +
(memorymap[i]*samples)]);
                          }
                    fprintf(data[i],"%4.0f\n",Fs);// write sampling
frequency
                    fprintf(data[i],
"%4i\n%2i\n%2i\n%2i\n%2i\n%2i\n",d.da_year,d.da_mon,d.da_day,t.ti_hour,t.
ti_min,t.ti_sec); // write date (year month day) and time (hour min sec)
                    if(fclose(data[i]))
                          {
                          printf("Error closing files");
                          return NO;
                          }
                    }
              }
        }
if(fclose(logfile))
        {
        printf("Error closing logfile");
        return NO;
        }

return 0K;

} // end of data2file


void wait(void)
{
int i,j;
for (i=0;i<Tconv;i++) j=i;
} /* end of wait */


void beep(void)
{
sound(1000); delay(50); nosound(); delay(100); sound(1000); delay(50);
nosound();
}
```

```
/* HARDWARE INTERRUPT FUNCTION ____________________________________________*/

void interrupt acquire()
{
disable();                                  // disable all IRQs
outportb(IMR,((inportb(IMR))|IRQ));         // mask IRQ
dummy = inportb(IOPORTA);              // read IOPORTA to reset IRQ flag

// pulse up IOPORTB bit 7 to signal ISR
outportb(IOPORTB,inportb(IOPORTB)|BIT7);

// CONVERSION

for(i=0;i<4;i++)
      {
      if (channellist[i] == OK)
            {
            outportb(IOPORTA2, inportb(IOPORTA2)|EN);//select analog mux
            outportb(IOPORTA2, inportb(IOPORTA2)&(~A0));     // reset A0
            outportb(IOPORTA2, inportb(IOPORTA2)&(~A1));     // reset A1
            mpxaddressmask = i;
            mpxaddressmask = mpxaddressmask << 1;
            mpxaddressmask = mpxaddressmask & 0x06;

            // select channel
            outportb(IOPORTA2, inportb(IOPORTA2)|mpxaddressmask);
            // deselect analog mux
            outportb(IOPORTA2, inportb(IOPORTA2)&(~EN));
            outportb(IOPORTA2, inportb(IOPORTA2)&(~HBEN));  // reset HBEN
            outportb(IOPORTA2, inportb(IOPORTA2)&(~CS));//select chip
            // read request -> conversion starts
            outportb(IOPORTA2, inportb(IOPORTA2)&(~RD));
            // wait untill conversion is finished
            while((inportb(IOPORTC2) & BUSY) != BUSY);
            // read lowerbyte D0-D7
            lowerbyte = inportb(IOPORTB2);
            outportb(IOPORTA2, inportb(IOPORTA2)|RD); // end read request
            outportb(IOPORTA2, inportb(IOPORTA2)|CS); // deselect chip
            // set HBEN -> redirect D8-D11 to D0-D3
            outportb(IOPORTA2, inportb(IOPORTA2)|HBEN);
            outportb(IOPORTA2, inportb(IOPORTA2)&(~CS));// select chip
            // read request -> conversion starts
            outportb(IOPORTA2, inportb(IOPORTA2)&(~RD));
            upperbyte = inportb(IOPORTB2);// read upperbyte D8-D11
            outportb(IOPORTA2, inportb(IOPORTA2)|RD); // end read request
            outportb(IOPORTA2, inportb(IOPORTA2)|CS); // deselect chip
            measure = lowerbyte + 256*upperbyte;

            // save measure2 to array
            datavect[index+(memorymap[i]*numberofsamples)] = measure;
            }
      }
index += 1;       // increment interruption counter

//pulse down bit7 to signal end of ISR
outportb(IOPORTB,inportb(IOPORTB)&(~BIT7));
outportb(IMR,((inportb(IMR))&(~IRQ)));          // demask IRQ
enable();               // re-enable all IRQs
outportb(EOI,EOI);// tell CPU that IRQ has been processed succesfully

} /* end of interrupt acquire */
```

```
// header file SYNACQ11.H

/* CONSTANTS DEFINITIONS for SYNACQ ------------------------------- */

// ADDA12 #1
#define ADDAPORT          0x02F0            // ADDA12 base address
#define ADPORTLOW         ADDAPORT+1        // AD lower byte
#define ADPORTHIGH        ADDAPORT+2        // AD upper byte
#define ADREGISTER        ADDAPORT+3        // AD register
#define ADPORTHIGHSWITCH ADDAPORT+4         // AD lowerbyte conversion
switch
#define ADPORTLOWSWITCH ADDAPORT+5          // AD upperbyte conversion
switch
#define DAPORTLOW         ADDAPORT+6        // DA lower byte
#define DAPORTHIGH        ADDAPORT+7        // DA upper byte

// ADDA12 #2
#define ADDAPORT2         0x0270            // ADDA12 base address2
#define ADPORTLOW2        ADDAPORT2+1       // AD lower byte
#define ADPORTHIGH2       ADDAPORT2+2       // AD upper byte
#define ADREGISTER2       ADDAPORT2+3       // AD register
#define ADPORTHIGHSWITCH2 ADDAPORT2+4       // AD lowerbyte conversion
switch
#define ADPORTLOWSWITCH2 ADDAPORT2+5        // AD upperbyte conversion
switch
#define DAPORTLOW2        ADDAPORT2+6       // DA lower byte
#define DAPORTHIGH2       ADDAPORT2+7       // DA upper byte

// ADC-42
#define ADC42             0x0300
#define ADCBUSYFLAG       ADC42 + 0
#define ADCRESULTHIGH     ADC42 + 1
#define ADCRESULTLOW      ADC42 + 2
#define ADCMPX            ADC42 + 0x000C

// 8259 Peripheral Interruption Controller
#define INTNUM     0x05  // Interrpution Request Number
#define IRQ        0x20  // mask for IRQ #5
#define IMR        0x21  // Interruption Mask Register for 8259 #1
#define IMR2       0x22  // Interruption Mask Register for 8259 #2
#define EOI        0x20  // hardware  End Of Interrupt

// 8255 chip #1
#define IOPORTA 0x01B0
#define IOPORTB IOPORTA+1
#define IOPORTC IOPORTA+2
#define IOCONTROLPORT IOPORTA+3

// 8255 chip #2
#define IOPORTA2 IOPORTA + 4
#define IOPORTB2 IOPORTA + 5
#define IOPORTC2 IOPORTA + 6
#define IOCONTROLPORT2 IOPORTA+ 7

// 8253
#define COUNTER0 IOPORTA + 8
#define COUNTER1 IOPORTA + 9
#define COUNTER2 IOPORTA + 10
#define COUNTERCONTROL  IOPORTA + 11

// 8253 Time Constants
#define Fc  1042000                  // board clock frequency in Hz
```

```
// 8253 Specific Modes
#define SQW0      0x36        // square wave generator on counter0
#define SHOT1     0x72        // one shoot mode on counter1 (diode pulse)
#define SHOT2     0xB2        // one shoot mode on counter2 (sample&hold)
#define RESET0    0x30        // reset counter0
#define RESET1    0x70        // reset counter1
#define RESET2    0xB0        // reset counter2

// 8255 Specific Modes
// (10011000b) mode 0 with porta input, portb output
#define MODE0AINBOUT 0x98

// (10111100b) mode 1 with porta input, portb output
#define MODE1AINBOUT 0xBC

// (10111111b) mode 1 with porta input, portb input
#define MODE1AINBIN  0xBF

// (10001010b) mode 0 with porta output, portb and portc input
#define MODE0AOUTBINCIN 0x8B

// (00001001b) enable  IOPORTC interrupt on IOPORTA !STB signal when
mode1 //active with IOPORTA as input
#define INTCE          0x09

// (00001000b) disable IOPORTC interrupt on IOPORTA !STB signal when
mode1 //active with IOPORTA as input
#define INTCD          0x08

// (00001001b) set bit C4 enable  IOPORTC interrupt on IOPORTA !STB
signal when //mode1 active with IOPORTA as input
#define INTAE          0x09

// (00001000b) reset bit C4 disable IOPORTC interrupt on IOPORTA !STB
signal //when mode1 active with IOPORTA as input
#define INTAD          0x08

// (00000101b) set bit C2 enable  IOPORTC interrupt on GROUP A IOPORTC
!STB //signal when mode1 with IOPORTA as input
#define INTBE          0x05

// (00000100b) reset bit C2 disable IOPORTC interrupt on GROUP B IOPORTC
!STB //signal when mode1 with IOPORTA as input
#define INTBD          0x04

// Amplifier Subtractor DCP Constants
#define DCPRESET        0x64  // d100 reset digital pot
#define AMP1            0x01  // b00000001 d1
#define AMP2            0x02  // b00000010 d2
#define SUB1            0x04  // b00000100 d4
#define SUB2            0x08  // b00001000 d8
#define UPDOWN          0x10  // b00010000 d16 up/!down mode
#define INC             0x20  // b00100000 d32 incremement wiper position
#define R1 10000.0// impedance of the DCP in use in the amplifier
#define R2 20000.0// impedance of the feedback resistor of the amplifier
#define MAXGAIN         R2/R1 // maximum voltage gain of the amplifier
#define SCALECONVERT    99.0/4095.0 // scale conversion factor
```

```
// Miscellanous constants
#define ADCHANNEL0      0x00        // AD conversion channel #0
#define ADCHANNEL1      0x01        // AD conversion channel #1
#define ADCHANNEL2      0x02        // AD conversion channel #2
#define ADCHANNEL3      0x03        // AD conversion channel #3
#define MAXCOUNT        400         // moving average period

#define Tconv     30                // AD conversion trigger delay
#define Pi        3.1415927
#define OK        0x01
#define ON        0x01
#define NO        0x00
#define OFF       0x00
#define BIT0      0x01
#define BIT1      0x02
#define BIT2      0x04
#define BIT3      0x08
#define BIT4      0x10
#define BIT5      0x20
#define BIT6      0x40
#define BIT7      0x80

// MAX 172  ADC
#define HBEN      0x20  // BIT5     on IOPORTA2
#define CS  0x08  // BIT3     on IOPORTA2
#define RD  0x10  // BIT4 on IOPORTA2
#define BUSY      0x01  // BIT0     on IOPORTC2

// DG409    analog multiplexer
#define A0  0x02  // BIT1 on IOPORTA2
#define A1  0x04  // BIT2 on IOPORTA2
#define EN  0x01  // BIT0 on IOPORTA2
```

APPENDIX 2 THE ZERO-PHASE FILTERING ALGORITHM

The zero-phase filtering algorithm is shown and detailed explanations of each step are given. The principle of the algorithm is to 1) filter a signal, then 2) reverse the filtered signal in time, then 3) filter the reversed signal and 4) revert the signal again. The result is equivalent to the effect of a single filter having an impulse response that is symmetrical in time with respect to the time origin. In other words, both the past and the future of a signal are taken into account during the convolution of the impulse response with the signal.

Suppose the linear and time invariant filter H has a real valued, causal impulse response h(t). The following convolution process defines the filter input-output relation, where x(t) and y(t) are the periodic input and output signals of finite duration T.

$$y(t) = \int_0^{T_h} h(\tau) \cdot x(t-\tau) \cdot d\tau \qquad \text{(Eq. 1)}$$

T_h is the duration of the impulse response. It can be can be either finite or infinite. In the case it is infinite, one must imagine that x(t) is padded with zeros outside the time range [0;T]. The convolution is more commonly expressed as:

$$y(t) = x(t) \otimes h(t) \qquad \text{(Eq. 2)}$$

In the Fourier domain, the same equation is written as follows.

$$Y(jw) = X(jw) \cdot H(jw) \qquad \text{(Eq. 3)}$$

where jw is the complex frequency variable.

Next, the signal y(t) is then reversed in time and its equivalent signal y'(t) is written according to:

$$y'(t) = y(T-t) \qquad \text{(Eq. 4)}$$

There is no direct equivalent expression in the Fourier domain so it is worked out from the Fourier integral:

$$Y'(jw) = \int_0^T y'(t) \cdot e^{-jwt} \cdot dt = \int_0^T y(T-t) \cdot e^{-jwt} \cdot dt$$

To simplify the above expression, we use a change of variable: The direction of the time axis is inverted and its origin is set to value T. The Fourier integral is now written as follows:

$$Y'(jw) = \int_0^T y(t) \cdot e^{+jwt} \cdot e^{-jwT} \cdot dt = e^{-jwT} \cdot \int_0^T y(t) \cdot e^{jwt} \cdot dt \qquad \text{(Eq. 5)}$$

The Fourier integral now corresponds to a Fourier transform with negative frequencies. Consequently, the expression for Y'(jw) is now written as:

$$Y'(jw) = e^{-jwT} \cdot Y(-jw) \qquad \text{(Eq. 6)}$$

The term Y(jw) is a sum of complex exponential terms. Therefore, a negative frequency argument will altern the sign of the imaginary part of the sum but the real part of Y(jw) will remain unaffected. Consequently a conjugate expression is produced. The expression above can be re-written as:

$$Y'(jw) = e^{-jwT} \cdot \overline{Y(jw)} \qquad \text{(Eq. 7)}$$

By expanding Y(jw) and manipulating the conjugated complex terms, we obtain:

$$Y'(jw) = e^{-jwT} \cdot \overline{X(jw) \cdot H(jw)} = e^{-jwT} \cdot \overline{X(jw)} \cdot \overline{H(jw)} \qquad \text{(Eq. 8)}$$

Next the reversed signal y'(t) is filtered again. In the time domain, this is written as:

$$z'(t) = y'(t) \otimes h(t) \qquad \text{(Eq. 9)}$$

where z'(t) is the output signal of the process. A Fourier domain expression is written:

$$Z'(jw) = Y'(jw) \cdot H(jw) = e^{-jwT} \cdot \overline{Y(jw)} \cdot H(jw) \qquad \text{(Eq. 10)}$$

Hence,

$$Z'(jw) = e^{-jwT} \cdot \overline{X(jw)} \cdot \overline{H(jw)} \cdot H(jw) \qquad \text{(Eq. 11)}$$

The product of H(jw) with its conjugated form returns a scalar expression. The expression above is re-written as follows:

$$Z'(jw) = e^{-jwT} \cdot \overline{X(jw)} \cdot |H(jw)|^2 \qquad \text{(Eq. 12)}$$

Finally, the filtered output time signal z'(t) is reversed in time. This is expressed as:

$$z(t) = z'(T - t) \qquad \text{(Eq. 13)}$$

Using equation 13 above, a equivalent expression in the Fourier domain is:

$$Z(jw) = e^{-jwT} \cdot \overline{Z'(jw)} \qquad \text{(Eq. 14)}$$

Finally, by expansion of Z'(jw) and elimination of the complex exponential terms, this becomes equivalent to:

$$Z(jw) = X(jw) \cdot |H(jw)|^2 \qquad \text{(Eq. 15)}$$

The square modulus of H(jw), termed $H_0(w)$, is the transmittance of the equivalent zero-phase filter. The complex variable j is voluntarily omitted in the frequency argument of H_0. A square modulus is no longer a complex expression and therefore carries no information regarding the phase difference between x(t) and z(t). As a result, z(t) is not only a filtered version of x(t) but also remains strictly in phase with x(t), independently of the type of impulse response used in the filter H. It is easily seen that the order of the zero-phase filter H_0 is twice the order of the original filter H.

We are now interested in determining the impulse response of the zero-phase filter H_0. An inverse Fourier transform of $H_0(jw)$ is expressed as:

$$h_o(t) = \int_{-\infty}^{+\infty} H_0(w) \cdot e^{+jwt} \cdot dw \qquad \text{(Eq. 16)}$$

Remembering that h(t) is real valued, it can be regarded as a sum of real sine and cosine terms. The Fourier transform of these sine and cosine terms have respectively an odd and even parity. Naturally H(jw) and $H_0(w)$ inherit this property and we can say that $H_0(w)$ equals to $H_0(-w)$. Consequently the equation above can be re-written as:

$$h_o(t) = \int_{0}^{+\infty} H_0(w) \cdot (e^{+jwt} + e^{-jwt}) \cdot dw = \int_{0}^{+\infty} H_0(w) \cdot 2\cos(wt) \cdot dw$$

We can see that the impulse response $h_0(t)$ is a sum of weighted cosine terms and therefore possesses an even parity with respect to time t = 0. This means that $h_0(t)$ extends in negative time. Convolution with such an impulse response corresponds to using both the past and the future of the signal being filtered, which is not causal. It is concluded that the effect of phase cancellation is due to the equal weighting of the signal by the impulse response around the present time.

If $h_0(t)$ has a finite duration, a causal version of the phase cancellation filter is obtained by delaying $h_0(t)$ by half its duration. In this case, the output signal looks as if it was obtained from a zero phase filter, but is delayed by half the duration of $h_0(t)$.
The zero phase filter has no causal (in the case of a real time implementation) equivalent for infinite duration impulse response.

APPENDIX 3 OFFLINE, ECG ASSISTED DETECTION OF ERRONEOUS FLAGS

```
%Variables preparation

Fs = 2000;                              %sampling frequency, 2000 Hz
Nstart = 2*Fs;                          %start point at 2 s
Nstop = 28 * Fs;                        %stop point at 28 s

ecg.r2r.next = zeros(N,1);          %next R2R interval
ecg.r2r.last = zeros(N,1);          %last R2R interval
tcg.max2max.next = zeros(N,1);      %next max2max interval
tcg.max2max.last = zeros(N,1);      %last max2max interval
tcg.min2min.next = zeros(N,1);      %next min2min interval
tcg.min2min.last = zeros(N,1);      %last min2min interval

%determine interval of contiguous ecg flags
for i = Nstart:Nstop
   if ecg.flag(i) == 1
      %search next flag
      for j = i+1:Nstop
        if ecg.flag(j) == 1
            ecg.r2r.next(i) = j-i;
            break
        end
      end
      %search last flag
      for j = i-1:-1:Nstart
        if ecg.flag(j) == 1
            ecg.r2r.last(i) = i-j;
            break
        end
      end
    end
end

%compare time interval of contiguous tcg max and ecg flags
match = 1; nomatch = 0;
lasttcginterval = nomatch;
nexttcginterval = nomatch;
```

```
%--------------- DISCARD ERRONEOUS MAX FLAGS ----------------

for i = Nstart:Nstop
   if tcg.flag(i) == 1

      %search for next ecg flag
       for m = i+1:1:Nstop
        if ecg.flag(m) == 1
           break
        end
      end

      %search for previous ecg flag
       for j = i-1:-1:Nstart
        if ecg.flag(j) == 1
           break
        end
      end

      %search for previous tcg flag
       for n = i-1:-1:Nstart
        if tcg.flag(n) == 1
           break
        end
      end

      %DISCARD FLAGS

      %if tcg flag is before ecg flag, then discard tcg flag
      if (m-i)<(i-j)
         tcg.flag(i) = 0.5;
       %if tcg flag is later than tolerance after ecg flag, then %discard tcg flag
      elseif (i-j)> 0.2*Fs
         tcg.flag(i) = 0.5;
      %if two consecutive flags are too close, discard the earliest
      elseif (i-n)<0.2*Fs
           tcg.flag(n) = 0.5;
      else

    %CHECK THAT INTERVALS BETWEEN VALID FLAGS ARE CONSISTENT WITH R2R %INTERVALS

       %search for last max2max interval and compare with last R2R %interval, break if
       match
       for k = i-1:-1:Nstart
           if tcg.flag(k) == 1
               if abs(ecg.r2r.last(j)-(i-k)) <= 0.2*Fs
                   lasttcginterval = match;
                   break
               else
                   lasttcginterval = nomatch;
               end
           end
       end

           %search for next max2max interval and compare with next RR %interval, break if
           match
          for k = i+1:1:Nstop
       if tcg.flag(k) == 1
           if abs(ecg.r2r.next(j) - (k-i)) <= 0.2*Fs
               nexttcginterval = match;
                  break
               else
               nexttcginterval = nomatch;
               end
           end
       end

      %take decision on tcg flag validity
       if ((lasttcginterval == nomatch) & (nexttcginterval == nomatch))
            tcg.flag(i) = 0.5;
        end

       end
   end
end
```

```
%-------------------- DISCARD ERRONEOUS MIN FLAGS ----------------------

for i = Nstart:Nstop
   if tcg.flag(i) == -1

      %search for next ecg flag
       for m = i+1:1:Nstop
        if ecg.flag(m) == 1
           break
        end
      end

      %search for previous ecg flag
       for j = i-1:-1:Nstart
        if ecg.flag(j) == 1
           break
        end
      end

      %search for previous tcg flag
       for n = i-1:-1:Nstart
        if tcg.flag(n) == -1
           break
        end
      end

    %DISCARD FLAGS (This tests are based on the fact that the ECG flag is anterior the
    %negative flags)

    %if next tcg flag in before next ecg flag, then discard next tcg flag
      if (n-j)> 0
         tcg.flag(i) = -0.5;
    %if positions of present tcg flag and next ecg flag closer than tolerance, discard tcg
    %flag
      elseif (m-i) < 0.2*Fs
         tcg.flag(i) = -0.5;
    %if positions of present tcg flag and last ecg flag are closer than tolerance, discard
    %tcg flag
      elseif (i-j) < 0.2*Fs
         tcg.flag(i) = -0.5;
      end

    end
end
```

APPENDIX 4 OFFLINE, ECG INDEPENDENT DETECTION OF ERRONEOUS FLAGS

```
%------------  DISCARD ERRONEOUS MIN FLAGS ------------------


%LAYER I ----  DISCARD MIN FLAGS IF TOO CLOSE FROM EACH OTHER

for i = Nstop:-1:Nstart
   if tcg.flag(i) == -1

      %search for previous max flag
        for k = i-1:-1:i-0.2*Fs
          if tcg.flag(k) == -1
             break
          end
      end

      %if two consecutive max flags are too close, discard the earliest
      if (i-k)<0.2*Fs
         tcg.flag(i) = -0.25;
      end
   end
end
```

```
%LAYER II ---- DISCARD MIN FLAGS WHICH DO NOT BELONG TO A PERIODIC CHAIN

match = 1; unmatch = 0;

%search for tcg min flag
for i = Nstart:Nstop
   if ((tcg.flag(i) == -1) & (tcg.std(i) < -0.5))
      %a flag has been found at i

      %BY DEFAULT, FLAG IS CONSIDERED WRONG
      status = unmatch;
      tcg.flag(i) = -0.5;

      %search for few last pcg min flags
      for k = i-1:-1:i-1.5*Fs
       if ((tcg.flag(k) <= -0.5) & (tcg.std(k) < -0.5))
            %a min tcg flag has been found at k, in the past of i

            if (i -k)>=0.6*Fs % 0.6 s is equivalent to 100 BPM, max heart rate in scanner
               %tcg.flag(k) is not too close from tcg.flag(i)

               %search for few next pcg min flags
               for q = i+1:1:i+1.5*Fs
                  if ((tcg.flag(q) <= -0.5) & (tcg.std(q) < -0.5))
                     %a min tcg flag has been found at q, in the future of i

                     if(q-i)>= 0.6*Fs %so that tcg.flag(q) is not too close tcg.flag(i)

                        if abs((q-i) - (i-k)) <= 0.2*Fs
                           %tcg.flag(i) in the middle of tcg.flag(k) and tcg.flag(q)
                           %It means tcg.flag(i) is part of a periodic sequence of 3
                         %consecutive min flags
                           status = match;

                           %CHECK IF MATCHING PAST FLAG WAS CARDIAC
                           if tcg.flag(k) <= -1.25

                              %full enable flag at i
                              tcg.flag(i) = -1.5;

                              %TO AVOID PARRALLEL CHAINS disable flags in between k and i
                              %THIS IS A RECURSIVE PROCESS, BUT IS ONLY SWITCHED ON IF NO
                              %RISK OF UNSTABILITY
                              for w = k+1:1:i-1
                                  if tcg.flag(w) <= -0.5
                                  tcg.flag(w) = -0.25;
                                  end
                              end

                          else
                           %MATCHING PAST FLAG WAS NOT CARDIAC
                           %CHECK IF THERE ARE ANY PREVIOUS CARDIAC FLAG IN PAST INTERVAL
                              for l = i-1:-1:k+1
                                  if tcg.flag(l) <= -1.25
                                  %disable flag at i
                                      tcg.flag(i) = -0.75;
                                  break
                                  else
                                  %enable flag at i
                                      tcg.flag(i) = -1.25;
                                 end
                              end

                           end

                           %stop scanning for previous/next flags
                           break

                        end
                     end
                     end
                  end
                  %first break point
                  if status == match
                     break
                  end
               end
            end
         end
      end
      %second break point
   end
end
```

```
%generate min2min intervals

tcg.min2min.next = zeros(N,1);
tcg.min2min.last = zeros(N,1);

for i = Nstart:Nstop
   if tcg.flag(i) <= -1.25
      %search next min flag
      for j = i+1:Nstop
       if tcg.flag(j) <= -1.25
            tcg.min2min.next(i) = j-i;
            break
         end
      end
      %search last min flag
      for j = i-1:-1:Nstart
       if tcg.flag(j) <= -1.25
            tcg.min2min.last(i) = i-j;
            break
         end
      end
    end
end


%------------ DISCARD ERRONEOUS MAX FLAGS ----------------------

tcg.max2max.next = zeros(N,1);
tcg.max2max.last = zeros(N,1);

lasttcginterval = unmatch;
nexttcginterval = unmatch;

%LAYER I ----------  DISCARD MAX FLAGS if too close from each others


for i = Nstart:Nstop
   if tcg.flag(i) == 1

      %search for previous max flag
       for k = i-1:-1:Nstart
           if tcg.flag(k) == 1
           break
           end
      end

      %if two consecutive max flags are too close, discard the earliest
      if (i-k)<0.2*Fs
           tcg.flag(k) = 0.25;
      end

   end
end
```

```
%LAYER II ----- DISCARD MAX FLAGS WHICH DO NOT BELONG TO A PERIODIC CHAIN
match = 1; unmatch = 0;

%search for tcg max flag
for i = Nstart:Nstop


   if ((tcg.flag(i) == 1) & (tcg.std(i) > 0.5))
      %a flag has been found at i

      %BY DEFAULT, FLAG IS CONSIDERED WRONG
      status = unmatch;
      tcg.flag(i) = 0.5;

      %search for few last tcg max flags
      for k = i-1:-1:i-1.5*Fs
        if ((tcg.flag(k) >= 0.5) & (tcg.std(k) > 0.5))
            %a min tcg flag has been found at k, in the past of i

            if (i-k)>=0.6*Fs % 0.6 s is equivalent to 100 BPM, max heart rate in scanner
               %tcg.flag(k) is not too close from tcg.flag(i)

               %search for few next pcg min flags
               for q = i+1:1:i+1.5*Fs
                  if ((tcg.flag(q) >= 0.5) & (tcg.std(q) > 0.5))
                     %a min tcg flag has been found at q, in the future of i

                     if(q-i)>= 0.6*Fs %so that tcg.flag(q) is not too close tcg.flag(i)

                        if abs((q-i) - (i-k)) <= 0.2*Fs
                           %tcg.flag(i) in the middle of tcg.flag(k) and tcg.flag(q)
                          %It means tcg.flag(i) is part of  %periodic sequence of 3
                          %consecutive min flags
                           status = match;

                           %CHECK IF MATCHING PAST FLAG WAS CARDIAC
                           if tcg.flag(k) >= 1.25

                              %full enable flag at i
                              tcg.flag(i) = 1.5;

                              %TO AVOID PARRALLEL CHAINS disable flags in between k and i
                              %THIS IS A RECURSIVE PROCESS, BUT IS ONLY SWITCHED ON IF NO
                              %RISK OF UNSTABILITY
                              for w = k+1:1:i-1
                                  if tcg.flag(w) >= 0.5
                                  tcg.flag(w) = 0.25;
                                  end
                              end
                           else
                              %MATCHING PAST FLAG WAS NOT CARDIAC CHECK IF THERE ARE ANY
                              %PREVIOUS CARDIAC FLAG IN PAST INTERVAL
                              for l = i-1:-1:k+1
                                  if tcg.flag(l) >= 1.25
                                  %disable flag at i
                                      tcg.flag(i) = 0.75;
                                  break
                                  else
                                  %enable flag at i
                                      tcg.flag(i) = 1.25;
                                 end
                              end

                           end

                           %stop scanning for previous/next flags
                           break

                        end
                     end
                     end
               end
               %first break point
               if status == match
                  break
               end
            end
         end
      end
      %second break point
   end end
```

APPENDIX 5 ONLINE, ECG ASSISTED DETECTION OF ERRONEOUS FLAGS

```
%_____________________________ FLAGGING MIN TCG _____________________________

%tresholding TCG
th = -1.5; tcg.tre = zeros(N,1); tcg.diff = zeros(N,1);
for i = 2:N
   if tcg.std(i) < th
      if tcg.std(i-1) < th
        tcg.tre(i) = th - tcg.std(i);
          tcg.diff(i) = tcg.tre(i) - tcg.tre(i-1);
      end
   end
end

%cubing 1st derivative of TCG. cube will compress anything below 1.
%Consequeltly, signal to noise ratio is increased and maxima detection should be easier.
tcg.flag = zeros(N,1); tcg.diff3 = zeros(N,1); tcg.diff3 = tcg.diff.^3;

%flagging minima
off = 0; on = 1; lock = off;
for i = 1:N
   if tcg.diff3(i) <= 0
      if lock == off
        tcg.flag(i) = -1;
            lock = on;
      end
   else
      lock = off;
    end
end


%_____________________________ FLAGGING MAX TCG _____________________________

%tresholding and flagging TCG max
th = 1.2; off = 0; on = 1; lock = off;
tcg.diff3 = zeros(N,1); tcg.tre = zeros(N,1); tcg.diff = zeros(N,1);
for i = 2:N
   tcg.diff(i) = tcg.std(i) - tcg.std(i-1); tcg.diff3(i) = tcg.diff(i).^3;
   if ((tcg.std(i)<=th)&(tcg.std(i)>0)&(tcgdiff3(i)<=0)&(tcg.diff3(i)>=-5E-9))
      if lock == off
        tcg.flag(i) = 1;
          lock = on;
      end
   else
    lock = off;
    end
end


%_____________________________ DISCARD ERRONEOUS TCG FLAGS _____________________________


ecg.r2r.next = zeros(N,1);
ecg.r2r.last = zeros(N,1);
tcg.max2max.next = zeros(N,1);
tcg.max2max.last = zeros(N,1);
tcg.min2min.next = zeros(N,1);
tcg.min2min.last = zeros(N,1);


%determine last ecg r2r interval
for i = Nstart:Nstop
   if ecg.flag(i) == 1
      %search last flag
      for j = i-1:-1:i-2*Fs
        if ecg.flag(j) == 1
            ecg.r2r.last(i) = i-j;
            break
          end
      end
    end
end
```

```
%compare time interval of contiguous tcg max and ecg flags
match = 1;
nomatch = 0;
lasttcginterval = nomatch;
nexttcginterval = nomatch;


%search for tcg min flag
for i = Nstart:Nstop
   if tcg.flag(i) == -1


      %search for previous ecg flag
       for j = i-1:-1:i-1.5*Fs
         if ecg.flag(j) == 1
            break
         end
      end

      %search for previous tcg flag
       for n = i-1:-1:i-1.5*Fs
         if tcg.flag(n) <= -0.5
            break
         end
      end


      %DISCARD ISOLATED FLAGS

      %if tcg flag is too early or too late after ecg flag, then discard tcg flag
      if (i-j)> 0.3*Fs, tcg.flag(i) = -0.25;
      elseif (i-j)< 0.1*Fs , tcg.flag(i) = -0.25;
         %if two consecutive tcg flags are too close, then discard tcg flag
      elseif (i-n)< 0.4*Fs, tcg.flag(i) = -0.5;
      else

         %DISCARD APERIODIC FLAGS BY FINDING PERIODIC FLAGS

       %search for last min2min interval and compare with last RR interval, break if match
       for k = i-1:-1:i-1.5*Fs
           if tcg.flag(k) <= -0.5
               if abs(ecg.r2r.last(j)-(i-k)) <= 0.2*Fs,
                  %FLAG at location i is very prob cardiac
                  tcg.flag(i) = -1.25;

                  %if last flag was cardiac , then this one is very probably cardiac too
                  if tcg.flag(k) <= -1.25, tcg.flag(i) = -1.5; end

                  %SCAN IN PAST 150 ms TO FIND A MAX
                          for w=i:-1:i-0.15*Fs,if tcg.flag(w)==1,tcg.flag(w)=1.5; break,
        end, end

                  %STOP SCANNING FOR PERIODICITY
                  break

               else
                  tcg.flag(i) = -0.5;
               end
           else
               %if there wasnt any flags in the last 1.5s, then nature of present flag is
               %uncertain
               tcg.flag(i) = -0.75;
           end
       end
         %break point
      end

   end
end
```

```
ecg.r2r.next = zeros(N,1);
ecg.r2r.last = zeros(N,1);

%determine interval of contiguous ecg flags
for i = Nstart:Nstop
   if ecg.flag(i) == 1
      %search next flag
      for j = i+1:Nstop
       if ecg.flag(j) == 1
            ecg.r2r.next(i) = j-i;
            break
         end
      end
      %search last flag
      for j = i-1:-1:Nstart
       if ecg.flag(j) == 1
            ecg.r2r.last(i) = i-j;
            break
         end
      end
    end
end


Nwin = 2*Fs;
Nheartbeatmax = 50;
Nheartbeat = 0;
last_n = 1;

win.ecg.norm = zeros(Nwin,Nheartbeatmax);
win.ecg.mean = zeros(Nwin,1);
win.ecg.r2r = zeros(1,Nheartbeatmax);

win.tcg.std = zeros(Nwin,Nheartbeatmax);
win.tcg.detrend = zeros(Nwin,Nheartbeatmax);
win.tcg.flag = zeros(Nwin,Nheartbeatmax);
win.tcg.detrend.bessel.real = zeros(Nwin,Nheartbeatmax);
win.tcg.detrend.bessel.offline = zeros(Nwin,Nheartbeatmax);
win.tcg.mean = zeros(Nwin,1);

for n = Nstart:Nstop,
    if(ecg.flag(n) == 1)
    Nheartbeat = Nheartbeat + 1;
      if Nheartbeat >= 2
         win.ecg.r2r(1,Nheartbeat) = ecg.r2r.next(n);
        win.ecg.norm(1:Nwin,Nheartbeat) = ecg.norm(n - Nwin/2:n + Nwin/2 - 1);
                win.tcg.detrend.bessel.real(1:Nwin,Nheartbeat) = tcg.detrend.bessel.real(n
        - Nwin/2:n + Nwin/2 - 1);
         win.tcg.detrend.bessel.offline(1:Nwin,Nheartbeat) = tcg.detrend.bessel.offline(n -
Nwin/2:n + Nwin/2 - 1);
         win.tcg.std(1:Nwin,Nheartbeat) = tcg.std(n - Nwin/2:n + Nwin/2 - 1);
         win.tcg.flag(1:Nwin,Nheartbeat) = tcg.flag(n - Nwin/2:n + Nwin/2 - 1);

       end
    end
end

% !!!! At that point, we have detected (Nheartbeat -1) number of heart beats

%clear empty colons at beginning
win.ecg.norm(:,1) = [];
win.ecg.r2r(:,1) = [];
win.tcg.std(:,1) = [];
win.tcg.flag(:,1) = [];
win.tcg.detrend.bessel.real(:,1) = [];
win.tcg.detrend.bessel.offline(:,1) = [];

%clear empty colons at end
win.ecg.norm(:,Nheartbeat:Nheartbeatmax-1) = [];
win.ecg.r2r(:,Nheartbeat:Nheartbeatmax-1) = [];
win.tcg.std(:,Nheartbeat:Nheartbeatmax-1) = [];
win.tcg.flag(:,Nheartbeat:Nheartbeatmax-1) = [];
win.tcg.detrend.bessel.real(:,Nheartbeat:Nheartbeatmax-1) = [];
win.tcg.detrend.bessel.offline(:,Nheartbeat:Nheartbeatmax-1) = [];

% get mean, diff of mean and SD of each row
win.ecg.mean(1:Nwin,1) = mean(win.ecg.norm(1:Nwin,:),2);
win.tcg.mean(1:Nwin,1) = mean(win.tcg.std(1:Nwin,:),2);
win.tcg.sd(1:Nwin,1) = std(win.tcg.std(1:Nwin,:),0,2);
win.tcg.dif(2:Nwin,1) = normalise(Fs*diff(win.tcg.mean(1:Nwin)));
```

```
% time ramp to display mean and std
win.t = [-Nwin/2:Nwin/2-1]'/Fs;

% convert interval to ms
win.ecg.r2r = win.ecg.r2r/Fs;


%_____________________ GENERATE R2MIN and R2MAX TIME INTERVALS_____________

Nheartbeat = 0;
Nheartbeatmax = size(win.tcg.std,2);
win.r2max = zeros(1,Nheartbeatmax);
win.r2min = zeros(1,Nheartbeatmax,1);

%detect max flag
for j = 1:Nheartbeatmax
    for i = Nwin/2:Nwin
        if win.tcg.flag(i,j) >= 1.25
            win.r2max(1,j) = i - Nwin/2;
            break
      end
   end
end

%detect min flag
for j = 1:Nheartbeatmax
    for i = Nwin/2:Nwin
        if win.tcg.flag(i,j) <= -1.25
        win.r2min(1,j) = i - Nwin/2;
          break
      end
    end
end

%convert intervals to milliseconds
win.r2max = win.r2max/Fs;
win.r2min = win.r2min/Fs;


%_____________________________________________ ECG.PERIOD _____________________

ecg.period = Fs*ones(N,1);

last_i = 1;
for i = 1:N
   if(ecg.flag(i)) == 1
    ecg.period(last_i:i) = i - last_i;
      last_i = i;
    end
end

ecg.period = ecg.period/Fs;

%_____________________________________________ TCG.PERIOD _____________________

tcg.period = Fs*ones(N,1);

last_i = 1;
for i = 1:N
   if tcg.flag(i) <= -1.25
      tcg.period(last_i:i) = i - last_i;
      last_i = i;
   end
end

tcg.period = tcg.period/Fs;
```

```
%_____ DOWNSAMPLING DATA to Fs/10 AND MODIFY NAMES ACCORDINGLY_____

%------  >>>>  downsampling windowed signals <<<<< ----------

ds_win_ecg_norm(:,:) = win.ecg.norm(1:10:Nwin,:);
ds_win_tcg_std(:,:) = win.tcg.std(1:10:Nwin,:);
ds_win_t(:,1) = win.t(1:10:Nwin,1);

%------- >>>> downsampling entire signal <<<<< ----------

ds_t(:,1) = t(1:10:N);
ds_tcg_std(:,1) = tcg.std(1:10:N);

%preprocessing flags before downsampling
tcg.dsflag = zeros(N,1); for i = 6:N-6, if tcg.flag(i) ~= 0, tcg.dsflag(i-5:i+5) =
tcg.flag(i); end, end
ecg.dsflag = zeros(N,1); for i = 6:N-6, if ecg.flag(i) ~= 0, ecg.dsflag(i-5:i+5) =
ecg.flag(i); end, end

ds_tcg_flag(:,1) = tcg.dsflag(1:10:N);
ds_ecg_flag(:,1) = ecg.dsflag(1:10:N);


ds_tcg_period(:,1) = tcg.period(1:10:N,1);
ds_ecg_period(:,1) = ecg.period(1:10:N,1);

win_r2max = win.r2max';
win_r2min = win.r2min';


%_____________________________________________________________________


%plot(t,ecg.flag,t,0.5*tcg.std,t,tcg.flag); grid
plot(win.t,win.tcg.std,win.t,win.tcg.flag); grid

%date and time
workspacename
datestr(now,0)
toc
```

```
%_____________________________ FLAGGING MIN TCG _____________________________

%tresholding TCG
th = -1.5; tcg.tre = zeros(N,1); tcg.diff = zeros(N,1);
for i = 2:N
   if tcg.std(i) < th
      if tcg.std(i-1) < th
        tcg.tre(i) = th - tcg.std(i);
          tcg.diff(i) = tcg.tre(i) - tcg.tre(i-1);
      end
   end
end

%cubing 1st derivative of TCG. cube will compress anything below 1.
%Consequeltly, signal to noise ratio is increased and maxima detection should be easier.
tcg.flag = zeros(N,1); tcg.diff3 = zeros(N,1); tcg.diff3 = tcg.diff.^3;

%flagging minima
off = 0; on = 1; lock = off;
for i = 1:N, if tcg.diff3(i) <= 0, if lock == off, tcg.flag(i) = -1; lock = on; end, else
lock = off; end, end

%_____________________________ FLAGGING MAX TCG _____________________________

%tresholding and flagging TCG max
th = 1.0; off = 0; on = 1; lock = off;
tcg.diff3 = zeros(N,1); tcg.tre = zeros(N,1); tcg.diff = zeros(N,1);
for i = 2:N
   tcg.diff(i) = tcg.std(i) - tcg.std(i-1); tcg.diff3(i) = tcg.diff(i).^3;
   if ((tcg.std(i) < th) & (tcg.std(i) > 0) & (tcg.diff3(i) <= 0) & (tcg.diff3(i) >= -5E-
9))
      if lock == off
        tcg.flag(i) = 1;
          lock = on;
      end
   else
    lock = off;
    end
end


%_____________________________ DISCARD ERRONEOUS TCG FLAGS _____________________


%The ecg r2r interval is not being used.
%It is being replaced by the previous tcg min2min interval.
%This is because they are fewer mins than max so good tcg min flags are easier to locate.


%_____________________ TCG MIN FLAGS _____________________________

%DISCARD APERIODIC TCG MIN FLAGS BY FINDING PERIODICS FLAGS.
%THIS PARTICULAR ALGO CANNOT HAVE SIMPLE STRUCTURE BECAUSE TWO PARAMETERS ARE TESTED
SIMULTANEOUSLY
%THIS IS WHY IT HAS GOT SO MANY NESTED LOOPS, WHICH MAKES DIFFICULT IT TO GRASP

match = 1;
unmatch = 0;

%error margin is called tolerance.
%200 ms is roughly the lenght of a peak at threshold level  = 1
tolerance = 0.2*Fs;

%search for tcg min flags
for i = Nstart:Nstop
   if tcg.flag(i) == -1
      %a flag has been found at i, now search for previous tcg flag
        for n = i-1:-1:i-1.25*Fs
          if tcg.flag(n) <= -0.5
             break
          end
      end
        %if two consecutive flags are too close, then discard present tcg flag
      if (i-n)< 0.2*Fs
         tcg.flag(i) = -0.25;
      else
           %NO FLAG IN THE last 200 ms of i SO WE START PERIODICITY ANALYSIS
```

```
        %BY DEFAULT, FLAG IS CONSIDERED WRONG ANYWAY
        status = unmatch;
        tcg.flag(i) = -0.5;
        %search for few last pcg min flags
        for k = i-1:-1:i-1.5*Fs    %
            if tcg.flag(k) <= -0.5
                %a min tcg flag has been found at k, in the past of i
                if (i -k)>= 0.6*Fs % 0.6 s is equivalent to 100 BPM,
                                   % max heart rate in scanner
                    %tcg.flag(k) is not too close from tcg.flag(i)
                    %search for last last pcg min flags
                    for q = k-1:-1:k-1.5*Fs
                    if tcg.flag(q) <= -0.5
                        %a min tcg flag has been found at q, in the future of i
                        if(k-q)>= 0.6*Fs
                        %tcg.flag(q) is not too close tcg.flag(i)
                                if abs((i-k) - (k-q)) <= 0.4*Fs %WE AGGRE THAT THE HEART
                                PERIOD CANNOT VARY BY MORE THAN 400 ms in one cycle
                                %tcg.flag(i) in the middle of tcg.flag(k) and tcg.flag(q)
                                %It means tcg.flag(i) is part of a periodic sequence of 3
                                consecutive min flags
                                status = match;
                                %CHECK IF MATCHING PAST FLAG WAS CARDIAC
                                if (tcg.flag(k) <= -1.25) & (tcg.flag(q) <= -1.25)
                                  %full enable flag at i
                                    tcg.flag(i) = -1.5; %TO AVOID PARRALLEL CHAINS disable
                                                        %flags in between k and i
                            %THIS IS A RECURSIVE PROCESS, BUT IS ONLY SWITCHED ON IF
                            %MINIMAL RISK OF PROPAGATING AN ERROR
                            %MINIMAL RISK IS ASSESSD BY nature of last two previous flags.
                                    for w=k+1:1:i-1,
                                        if tcg.flag(w)<=-0.5,
                                            tcg.flag(w)= -0.25;
                                        end,
                                    end
                                    for w = q+1:1:k-1,
                                        if tcg.flag(w) <= -0.5,
                                            tcg.flag(w) = -0.25;
                                        end,
                                    end
                                    %SCAN IN PAST 150 ms TO FIND A MAX
                                    for w = i:-1:i - 0.150*Fs,
                                        if tcg.flag(w) == 1,
                                            tcg.flag(w) = 1.5; break,
                                        end,
                                    end
                                else
                              %MATCHING PAST FLAGS WERE NOT CARDIAC.
                              %NOW CHECK IF THERE ARE ANY PREVIOUS CARDIAC FLAG IN PAST INTERVAL
                                        for l = i-1:-1:k+1,
                                           if tcg.flag(l) == -1.25,
                                              tcg.flag(i) = -0.5;
                                              break
                                           else
                                              tcg.flag(i) = -1.25;
                                              %SCAN IN PAST 150 ms TO FIND A MAX
                                             for w = i:-1:i - 0.15*Fs,
                                                 if tcg.flag(w) == 1,
                                                     tcg.flag(w) = 1.25;
                                                         break,
                                                 end,
                                             end
                                           end
                                        end
                                    end
                                    %stop scanning for last flags
                                    break
                                end
                            end
                            end
                        end
                        %first break point
                        if status == match
                        break
                        end
                    end
                end
              end
            end
            %second break point
        end
    end
end
```

```
%generate min2min intervals

tcg.min2min.next = zeros(N,1);
tcg.min2min.last = zeros(N,1);

for i = Nstart:Nstop
   if tcg.flag(i) <= -1.25
      %search next min flag
      for j = i+1:Nstop
         if tcg.flag(j) <= -1.25
            tcg.min2min.next(i) = j-i;
            break
         end
      end
      %search last min flag
      for j = i-1:-1:Nstart
        if tcg.flag(j) <= -1.25
            tcg.min2min.last(i) = i-j;
            break
         end
      end
   end
end


%_____________________________________ WINDOWs ECG TCG ____________________________________

ecg.r2r.next = zeros(N,1);
ecg.r2r.last = zeros(N,1);

%determine interval of contiguous ecg flags
for i = Nstart:Nstop
   if ecg.flag(i) == 1
      %search next flag
      for j = i+1:Nstop
        if ecg.flag(j) == 1
            ecg.r2r.next(i) = j-i;
            break
         end
      end
      %search last flag
      for j = i-1:-1:Nstart
        if ecg.flag(j) == 1
            ecg.r2r.last(i) = i-j;
            break
         end
      end
   end
end


Nwin = 2*Fs;
Nheartbeatmax = 50;
Nheartbeat = 0;
last_n = 1;

win.ecg.norm = zeros(Nwin,Nheartbeatmax);
win.ecg.mean = zeros(Nwin,1);
win.ecg.r2r = zeros(1,Nheartbeatmax);

win.tcg.std = zeros(Nwin,Nheartbeatmax);
win.tcg.detrend = zeros(Nwin,Nheartbeatmax);
win.tcg.flag = zeros(Nwin,Nheartbeatmax);
win.tcg.detrend.bessel.real = zeros(Nwin,Nheartbeatmax);
win.tcg.detrend.bessel.offline = zeros(Nwin,Nheartbeatmax);
win.tcg.mean = zeros(Nwin,1);

for n = Nstart:Nstop,
   if(ecg.flag(n) == 1)
   Nheartbeat = Nheartbeat + 1;
     if Nheartbeat >= 2
        win.ecg.r2r(1,Nheartbeat) = ecg.r2r.next(n);
       win.ecg.norm(1:Nwin,Nheartbeat) = ecg.norm(n - Nwin/2:n + Nwin/2 - 1);
        win.tcg.detrend.bessel.real(1:Nwin,Nheartbeat) = tcg.detrend.bessel.real(n -
Nwin/2:n + Nwin/2 - 1);
        win.tcg.detrend.bessel.offline(1:Nwin,Nheartbeat) = tcg.detrend.bessel.offline(n -
Nwin/2:n + Nwin/2 - 1);
        win.tcg.std(1:Nwin,Nheartbeat) = tcg.std(n - Nwin/2:n + Nwin/2 - 1);
        win.tcg.flag(1:Nwin,Nheartbeat) = tcg.flag(n - Nwin/2:n + Nwin/2 - 1);
```

```
        end
    end
end

% !!!! At that point, we have detected (Nheartbeat -1) number of heart beats

%clear empty colons at beginning
win.ecg.norm(:,1) = [];
win.ecg.r2r(:,1) = [];
win.tcg.std(:,1) = [];
win.tcg.flag(:,1) = [];
win.tcg.detrend.bessel.real(:,1) = [];
win.tcg.detrend.bessel.offline(:,1) = [];

%clear empty colons at end
win.ecg.norm(:,Nheartbeat:Nheartbeatmax-1) = [];
win.ecg.r2r(:,Nheartbeat:Nheartbeatmax-1) = [];
win.tcg.std(:,Nheartbeat:Nheartbeatmax-1) = [];
win.tcg.flag(:,Nheartbeat:Nheartbeatmax-1) = [];
win.tcg.detrend.bessel.real(:,Nheartbeat:Nheartbeatmax-1) = [];
win.tcg.detrend.bessel.offline(:,Nheartbeat:Nheartbeatmax-1) = [];

% get mean, diff of mean and SD of each row
win.ecg.mean(1:Nwin,1) = mean(win.ecg.norm(1:Nwin,:),2);
win.tcg.mean(1:Nwin,1) = mean(win.tcg.detrend.bessel.real(1:Nwin,:),2);
win.tcg.sd(1:Nwin,1) = std(win.tcg.detrend.bessel.real(1:Nwin,:),0,2);
win.tcg.dif(2:Nwin,1) = diff(win.tcg.mean(1:Nwin));


% time ramp to display mean and std
win.t = [-Nwin/2:Nwin/2-1]'/Fs;

% convert interval to ms
win.ecg.r2r = win.ecg.r2r/Fs;


%_____ GENERATE R2MIN and R2MAX TIME _____

Nheartbeat = 0;
Nheartbeatmax = size(win.tcg.std,2);
win.r2max = zeros(1,Nheartbeatmax);
win.r2min = zeros(1,Nheartbeatmax,1);

%detect max flag
for j = 1:Nheartbeatmax
    for i = Nwin/2:Nwin
        if win.tcg.flag(i,j) >= 1.5
            win.r2max(1,j) = i - Nwin/2;
            break
      end
   end
end

%detect min flag
for j = 1:Nheartbeatmax
    for i = Nwin/2:Nwin
        if win.tcg.flag(i,j) <= -1.5
        win.r2min(1,j) = i - Nwin/2;
          break
      end
    end
end

%convert intervals to milliseconds
win.r2max = win.r2max/Fs;
win.r2min = win.r2min/Fs;


%_____ ECG.PERIOD _____

ecg.period = Fs*ones(N,1);

last_i = 1;
for i = 1:N
   if(ecg.flag(i)) == 1
    ecg.period(last_i:i) = i - last_i;
      last_i = i;
    end
end

ecg.period = ecg.period/Fs;
```

```
tcg.period = Fs*ones(N,1);

last_i = 1;
for i = 1:N
   if tcg.flag(i) >= 1.25
      tcg.period(last_i:i) = i - last_i;
      last_i = i;
   end
end

tcg.period = tcg.period/Fs;
```

APPENDIX 7 MULTIPLE LINEAR REGRESSIVE CALIBRATION

```
tic
clear
clc

load forp52

Fs = 2000;
N = length(t);

%-------------------- abdomen
A = -A; AA = A + 4095; A = A - mean(A);
b = ones(0.1*Fs,1); a = 0.1*Fs;
A = filtfilt(b,a,A);
A2 = A.^2;
A3 = A.^3;
A4 = A.^4;

%--------------------- ribcage
R = -R; RR = R + 4095; R = R - mean(R);
R = filtfilt(b,a,R);
R2 = R.^2;
R3 = R.^3;
R4 = R.^4;

%-------------------- spirometer
S = filtfilt(b,a,S);

%--------------------- ordre maximum du polynome de l'equation FORP
A_max = 2;
R_max = 2;

%------------------------>>>>>>>>>> building sigma values <<<<<<<<<<<<<<
S_sigma(1) = sum(S);

for i = 1:2*A_max
   A_sigma(i) = sum(A.^i);
   SA_sigma(i) = sum(S.*(A.^i));
end

for i = 1:2*R_max
   R_sigma(i) = sum(R.^i);
   SR_sigma(i) = sum(S.*(R.^i));
end

for i = 1:A_max,
   for j = 1:R_max,
      AR_sigma(i,j) = sum((A.^i).*(R.^j));
   end
end

%---------------------- >>>>>>>>>>>> linear regression <<<<<<<<<<<<<<<<

Alpha = [
   R_sigma(2)          AR_sigma(1,1)      R_sigma(1)
    AR_sigma(1,1)      A_sigma(2)         A_sigma(1)
   R_sigma(1)          A_sigma(1)         N
];

Beta = [
   SR_sigma(1)
   SA_sigma(1)
   S_sigma(1)
];

%calibration constants
K = inv(Alpha)*Beta; K = K';

%generate vitual respiration trace
F = K(1)*R + K(2)*A + K(3);

E = S - F;
```

```
%mean square of normalised error
offset = min(S);
s = S - offset; f = F - offset;
gain  = max(s);
s = s/gain; f = f/gain;
e = s - f;
MSE = mean(e.^2);

subplot(2,1,1); plot(t,S,t,F,t,R,t,A); grid
subplot(2,1,2); plot(t,E); grid

%---------- >>>>>>>>>>> downsampling to 200 Hz <<<<<<<<<<<

t = t';
ds_t(:,1) = t(1:10:N);
ds_R(:,1) = R(1:10:N);
ds_RR(:,1) = RR(1:10:N);
ds_A(:,1) = A(1:10:N);
ds_AA(:,1) = AA(1:10:N);
ds_S(:,1) = S(1:10:N);
ds_F(:,1) = F(1:10:N);

toc
```

APPENDIX 8 MULTIPLE QUADRATIC REGRESSIVE CALIBRATION

```
tic
clc
clear

load forp52

Fs = 2000;
N = length(t);

%-------------------- abdomen
A = -A; AA = A + 4095; A = A - mean(A);
b = ones(0.1*Fs,1); a = 0.1*Fs;
A = filtfilt(b,a,A);
A2 = A.^2;
A3 = A.^3;
A4 = A.^4;

%-------------------- ribcage
R = -R; RR = R + 4095; R = R - mean(R);
R = filtfilt(b,a,R);
R2 = R.^2;
R3 = R.^3;
R4 = R.^4;

%-------------------- spirometer
S = filtfilt(b,a,S);

%-------------------- ordre maximum du polynome de l'equation FORP
A_max = 2;
R_max = 2;


%------------------------>>>>>>>>>> building sigma values <<<<<<<<<<<<<<
S_sigma(1) = sum(S);

for i = 1:2*A_max
   A_sigma(i) = sum(A.^i);
   SA_sigma(i) = sum(S.*(A.^i));
end

for i = 1:2*R_max
   R_sigma(i) = sum(R.^i);
   SR_sigma(i) = sum(S.*(R.^i));
end

for i = 1:A_max,
   for j = 1:R_max,
      AR_sigma(i,j) = sum((A.^i).*(R.^j));
   end
end

%---------------------- >>>>>>>>>>>> linear regression <<<<<<<<<<<<<<<<

Alpha = [
   R_sigma(2)      R_sigma(3)      AR_sigma(1,1)      AR_sigma(2,1)      R_sigma(1)
   R_sigma(3)      R_sigma(4)      AR_sigma(1,2)      AR_sigma(2,2)      R_sigma(2)
   AR_sigma(1,1)       AR_sigma(1,2)  A_sigma(2)          A_sigma(3)         A_sigma(1)
   AR_sigma(2,1)       AR_sigma(2,2)  A_sigma(3)          A_sigma(4)         A_sigma(2)
   R_sigma(1)      R_sigma(2)      A_sigma(1)         A_sigma(2)         N

];

Beta = [
   SR_sigma(1)
   SR_sigma(2)
   SA_sigma(1)
   SA_sigma(2)
   S_sigma(1)
];

%calibration constants
K = inv(Alpha)*Beta; K = K';

%generate vitual respiration trace
F = K(1)*R + K(2)*R2 + K(3)*A + K(4)*A2 + K(5);
```

```
%generate error
E = S - F;

%mean square of normalised error
offset = min(S);
s = S - offset; f = F - offset;
gain  = max(s);
s = s/gain; f = f/gain;
e = s - f;
MSE = mean(e.^2);

subplot(2,1,1); plot(t,S,t,F,t,RR,t,AA); grid
subplot(2,1,2); plot(t,K(1)*R,t,K(3)*A,t,K(2)*R2,t,K(4)*A2); grid


%------------------------ >>>>>>>>>>> downsampling to 200 Hz <<<<<<<<<<<
t = t';
ds_t(:,1) = t(1:10:N);
ds_R(:,1) = R(1:10:N);
ds_A(:,1) = A(1:10:N);
ds_RR(:,1) = RR(1:10:N);
ds_AA(:,1) = AA(1:10:N);
ds_S(:,1) = S(1:10:N);
ds_F(:,1) = F(1:10:N);


toc
```

APPENDIX 9 WIENER LINEAR COMBINER CALIBRATION

```
tic
clc
close
close
clear
load forp52

Fs = 2000;
N = length(t);

t = t';
ds_t(:,1) = t(1:10:N);

%-------------------- abdomen ----------
A = -A; AA = A + 4095; A = A - mean(A);
b = ones(0.1*Fs,1); a = 0.1*Fs;
A = filtfilt(b,a,A);
y(:,1) = A(1:10:N);

%--------------------- ribcage ------------
R = -R; RR = R + 4095; R = R - mean(R);
R = filtfilt(b,a,R);
x(:,1) = R(1:10:N);

%-------------------- spirometer ---------------
p0 = mean(S);
S = S - mean(S);
S = filtfilt(b,a,S);
z(:,1) = S(1:10:N);


Fs = 200;
N = length(x);

%Wiener filter size
Nf = 1*Fs;


%z
%-----------------------------------------------
%(reference signal)                              |
%                                                |
%x                              xn               |  e
%--------------------> filter ----> * -1 ---> + ----------------
%(dirty signal)                                  |      (error signal)
%                                                |
%y                              yn               |
%--------------------> filter ----> * -1 -------|
%(dirty signal)

%correlation matrix
Rxx_vec = xcorr(x,x,Nf,'none');
Rxy_vec = xcorr(x,y,Nf,'none');
Ryx_vec = xcorr(y,x,Nf,'none');
Ryy_vec = xcorr(y,y,Nf,'none');

for i = 1:Nf,
   Rxx(:,i) = Rxx_vec(Nf+i:-1:i+1);
   Rxy(:,i) = Rxy_vec(Nf+i:-1:i+1);
   Ryx(:,i) = Ryx_vec(Nf+i:-1:i+1);
   Ryy(:,i) = Ryy_vec(Nf+i:-1:i+1);
end

Rmatrix = [Rxx     Ryx
           Rxy     Ryy];

%correlation vector
Pzx = xcorr(z,x,Nf,'none');
Pzx(1:Nf) = Pzx(Nf+1:-1:1+1);
Pzx(Nf+1:2*Nf+1) = [];

Pzy = xcorr(z,y,Nf,'none');
Pzy(1:Nf) = Pzy(Nf+1:-1:1+1);
Pzy(Nf+1:2*Nf+1) = [];

Pvec = [Pzx;   Pzy];
```

```
%calculating optimal filter coefficients with Wiener-Hopf equation
C = inv(Rmatrix)*Pvec;

%separating filter coefficents
Hx = C(1:Nf);
Hy = C(Nf+1:2*Nf);

%filtering
xn = filter(Hx,1,x);
yn = filter(Hy,1,y);

%generating estimation of f
f = xn + yn;

%get filters complex transfert function
[Tx,faxis] = freqz(Hx,1,N,Fs);
[Ty,faxis] = freqz(Hy,1,N,Fs);

%get transfert function module
modx = abs(Tx);
mody = abs(Ty);

%get phase
phix= unwrap(angle(Tx));
phiy= unwrap(angle(Ty));

%get phase delay
deltax = phix./(2*pi*faxis);
deltay = phiy./(2*pi*faxis);

K(1) = modx(2);
K(2) = deltax(2);
K(3) = mody(2);
K(4) = deltay(2);
K(5) = p0;

%------------------>>>>>>> error calculation <<<<<<<<----------

%mean square of normalised error
offset = min(z);
Z = z - offset; F = f - offset;
gain  = max(Z);
Z = Z/gain; F = F/gain;
E = Z - F;
MSE = mean(E.^2);


z = z + p0;
f = f + p0;
subplot(2,1,1); plot(ds_t,z,ds_t,f); grid
subplot(2,1,2); plot(ds_t,x,ds_t,xn,ds_t,y,ds_t,yn); grid

figure
subplot(2,1,1); semilogx(faxis,modx,faxis,mody); grid
subplot(2,1,2); semilogx(faxis,deltax,faxis,deltay); grid

toc
```

APPENDIX 10 LISTING OF THE TENSIOMETER CONTROL APPLICATION IN VB5

frmMain

```
Option Explicit

Dim Length As Double
Dim Position As Double
Dim InitialPosition As Double
Dim Run As Boolean
Dim RetVal
Dim Argument As String
Dim Status As Byte

Private Sub cmdDown_Click()

'set focus to stop button
cmdStop.SetFocus

'filter speed text box
If (txtSpeedManual.Text = "") Or (txtSpeedManual.Text < 10) Then
   txtSpeedSingle.Text = 10
End If

'filter speed text box
If txtSpeedManual.Text > 1000 Then txtSpeedManual.Text = 1000


'send speed to tensiometer
WriteMachine ("WV" + txtSpeedManual.Text)

'setting direction
WriteMachine ("WR")

'set run flag
Run = True

Do While Run = True
   GetPosition
Loop

End Sub


Private Sub cmdExit_Click()

'switch off test diode
WriteMachine ("WG0")

'Stop the machine
WriteMachine ("WS")

'Disconnect remote mode
WriteMachine ("WC")

End
```

```
End Sub


Private Sub cmdStart_Click()

If optSingle.Value = True Then
   SingleRun
ElseIf optReturn.Value = True Then
   ReturnRun
End If

End Sub



Private Sub cmdStop_Click()
Run = False
WriteMachine ("WS")

End Sub

Private Sub cmdUp_Click()

'set focus to stop button
cmdStop.SetFocus

'filter speed text box
If (txtSpeedManual.Text = "") Or (txtSpeedManual.Text < 10) Then
   txtSpeedSingle.Text = 10
End If

'filter speed text box
If txtSpeedManual.Text > 1000 Then txtSpeedManual.Text = 1000


'send speed to tensiometer
WriteMachine ("WV" + txtSpeedManual.Text)

'set run flag
Run = True

'setting direction
WriteMachine ("WF")


Do While Run = True
   GetPosition
Loop

End Sub

Private Sub Form_Load()

Anjan   'unlocking vbio.dll

frmMain.msComm1.PortOpen = True        'open port

'initialise text box
txtTimeSingle.Text = Str(60 * Val(txtLengthSingle.Text) / Val(txtSpeedSingle.Text))
txtTimeReturn.Text = Str(Val(txtPauseReturn.Text) + (120 * Val(txtLengthReturn.Text) /
```

```
Val(txtSpeedReturn.Text)))

'enable manual frame
fraSingle.Enabled = False
fraReturn.Enabled = False
fraManual.Enabled = False

'reset run flag
Run = False

'switch on test diode
WriteMachine ("WG1")

tmrWait.Enabled = False

'set 8255 to mode1 with IOPORTA2 as input and IOPORTB2 as output
Out IOCONTROLPORT2, MODE1AINBOUT

'reset ioportb2
Out IOPORTB2, 0

End Sub


Private Sub optManual_Click()
fraSingle.Enabled = False
fraReturn.Enabled = False
fraManual.Enabled = True

optSingle.BackColor = &H8000000F
optReturn.BackColor = &H8000000F
optManual.BackColor = vbBlue


End Sub

Private Sub optReturn_Click()
fraSingle.Enabled = False
fraReturn.Enabled = True
fraManual.Enabled = False

optSingle.BackColor = &H8000000F
optReturn.BackColor = vbBlue
optManual.BackColor = &H8000000F

End Sub


Private Sub optSingle_Click()
fraSingle.Enabled = True
fraReturn.Enabled = False
fraManual.Enabled = False


optSingle.BackColor = vbBlue
optReturn.BackColor = &H8000000F
optManual.BackColor = &H8000000F


End Sub
```

```
Public Sub SingleRun()

'set focus to stop button
cmdStop.SetFocus

'warn that text box was empty
If txtLengthSingle.Text = "" Then
   txtLengthSingle.Text = "0"
   Exit Sub
End If

'get run length
Length = Val(txtLengthSingle.Text)

'make sure text box is not empty
If txtSpeedSingle.Text = "" Then txtSpeedSingle.Text = 10


'make sure text box speed is not under 10
If txtSpeedSingle.Text < 10 Then txtSpeedSingle.Text = 10


'make sure text box speed is not above 1000
If txtSpeedSingle.Text > 1000 Then txtSpeedSingle.Text = 1000

'send speed to tensiometer
WriteMachine ("WV" + txtSpeedSingle.Text)

'set run flag
Run = True

'get initial position
InitialPosition = GetPosition

If ckcControlAcq.Value = 1 Then

   'set "tensiometer ready" flag
   Status = Inp(IOPORTA2) And HByteMask
   Status = Status Or BIT1
   Out IOPORTB2, Status

   'call SYNACQ NOW and pass Fs, duration , Speed
   Argument = "d:\tcpp\synacq17" + " " + txtTimeSingle.Text + " " + txtFs.Text + " " +
Str(Val(txtSpeedSingle.Text) / 60)
   RetVal = Shell(Argument, vbMinimizedNoFocus)
   If RetVal = 0 Then Exit Sub

   'test "synacq ready" flag
   Do                    'wait untill "tensiometer ready" flag set
      Status = Inp(IOPORTA2) And HByteMask
      If Status And BIT0 Then Exit Do
      'BIT0 has been set by SYNACQ , acquisition will now start
   Loop

End If

'setting direction
If optUpSingle.Value = True Then
   WriteMachine ("WF")
```

```
ElseIf optDownSingle.Value = True Then
   WriteMachine ("WR")
End If

Do
   GetPosition

   If Abs(Position - InitialPosition) > Length Then
      'stop tensiometer
      WriteMachine ("WS")
      'reset run flag
      Run = False
   End If

Loop Until Run = False

End Sub

Public Sub ReturnRun()

'set focus to stop button
cmdStop.SetFocus

'filter length text box
If txtLengthReturn.Text = "" Then
   txtLengthReturn.Text = "0"
   Exit Sub
End If

'get run length
Length = Val(txtLengthReturn.Text)

'filter speed text box is OK
If (txtSpeedReturn.Text = "") Or (txtSpeedReturn.Text < 10) Then
   txtSpeedSingle.Text = "10"
End If

'filter speed text box is OK
If txtSpeedReturn.Text > 1000 Then
   txtSpeedSingle.Text = "1000"
End If

'set tensiometer speed
WriteMachine ("WV" + txtSpeedReturn.Text)

'get Initial Position
InitialPosition = GetPosition

'set run flag
Run = True

If ckcControlAcq.Value = 1 Then

   'set "tensiometer ready" flag
   Status = Inp(IOPORTA2) And HByteMask
   Status = Status Or BIT1
   Out IOPORTB2, Status

   'call SYNACQ and pass Fs, duration , Speed
   Argument = "d:\tcpp\synacq17" + " " + txtTimeReturn.Text + " " + txtFs.Text + " " +
```

```
Str(Val(txtSpeedReturn.Text) / 60)
   RetVal = Shell(Argument, vbMinimizedNoFocus)

   'test "synacq ready" flag, start when set
   Do
      Status = Inp(IOPORTA2) And HByteMask
      If Status And BIT0 Then Exit Do
      'BIT0 has been set by SYNACQ , acquisition will now start
   Loop

End If

'set tensiometer direction for A->B run
If optUDUReturn.Value = True Then
   WriteMachine ("WF")
ElseIf optDUDReturn.Value = True Then
   WriteMachine ("WR")
End If


Do
   Position = GetPosition

   If Abs(Position - InitialPosition) > Length Then
      'stop tensiometer
      WriteMachine ("WS")
      'reset run flag
      Run = False
   End If

Loop Until Run = False

'wait
If Val(txtPauseReturn.Text) > 0 Then
   tmrWait.Interval = 1000 * Val(txtPauseReturn.Text)
   tmrWait.Enabled = True

   Do
      DoEvents
   Loop Until tmrWait.Enabled = False

End If


'get Initial Position
InitialPosition = Position


'set tensiometer direction for B->A run
If optUDUReturn.Value = True Then
   WriteMachine ("WR")
ElseIf optDUDReturn.Value = True Then
   WriteMachine ("WF")
End If

'set run flag
Run = True


Do
```

```
    GetPosition

    If Abs(Position - InitialPosition) > Length Then
       'stop tensiometer
       WriteMachine ("WS")
       'reset run flag
       Run = False
    End If

Loop Until Run = False


End Sub

Public Function GetPosition()
Position = Val(ReadMachine("RP")) / 1000
'frmMain.Caption = "Tensiometer Control                                    Position : " + Str(Position) +
" mm"
GetPosition = Position
End Function

Private Sub tmrWait_Timer()
   tmrWait.Enabled = False
End Sub

Private Sub txtLengthReturn_Change()
If Val(txtSpeedReturn.Text) <> 0 Then
   txtTimeReturn.Text = Str(Val(txtPauseReturn.Text) + 2 * (Val(txtLengthReturn.Text) /
(Val(txtSpeedReturn.Text) / 60)))
End If
End Sub

Private Sub txtLengthSingle_Change()
If Val(txtSpeedSingle.Text) <> 0 Then
   txtTimeSingle.Text = Str(60 * Val(txtLengthSingle.Text) / Val(txtSpeedSingle.Text))
End If
End Sub

Private Sub txtPauseReturn_Change()
If Val(txtSpeedReturn.Text) <> 0 Then
   txtTimeReturn.Text = Str(Val(txtPauseReturn.Text) + 2 * (Val(txtLengthReturn.Text) /
(Val(txtSpeedReturn.Text) / 60)))
End If
End Sub


Private Sub txtSpeedReturn_Change()
If Val(txtSpeedReturn.Text) <> 0 Then
   txtTimeReturn.Text = Str(Val(txtPauseReturn.Text) + 2 * (Val(txtLengthReturn.Text) /
(Val(txtSpeedReturn.Text) / 60)))
End If
End Sub

Private Sub txtSpeedSingle_Change()
If Val(txtSpeedSingle.Text) <> 0 Then
   txtTimeSingle.Text = Str(60 * Val(txtLengthSingle.Text) / Val(txtSpeedSingle.Text))
End If
End Sub
```

BLANK PAGE
IN
ORIGINAL

ModModule1

```
Public nAuxChannels(9) As Integer
Public Function WriteMachine(sCommand As String)

' Used for write commands such as WS, WF, WR, etc.
' Returns 1 for success or 0 for fail

Dim sLetter As String
Dim sBuffer As String
Dim nStart As Single
Dim nFinish As Single

 ' Open com port if not open
 If frmMain.msComm1.PortOpen = False Then frmMain.msComm1.PortOpen = True
 ' Send command
 frmMain.msComm1.Output = sCommand & Chr$(13)

 ' Set timeout variables
 nStart = Timer
 nFinish = nStart + 2

 Do
  DoEvents
  ' Get character from input buffer
  sLetter = frmMain.msComm1.Input
  ' If character is a carraiage return then response has been received
  If sLetter = vbCr Then
   ' If ? is return then command was not recognised
   If InStr(1, sBuffer, "?") Then
    ' Send command failed
    WriteMachine = 0
   Else
    ' Send command successful
    WriteMachine = 1
   End If
   ' Exit loop
   Exit Do
  Else
   ' If character is not a carriage return then append it to sBuffer
   sBuffer = sBuffer & sLetter
  End If
 ' Loop until timeout or response has been found
 Loop Until Timer >= nFinish Or bExit

End Function

Function ReadMachine(sCommand As String) As String

' Used for read commands such as RL, RP, RI, etc
' Returns result string for success or "" for fail

Dim iCounter As Integer
Dim iAuxChannel As Integer
Dim lIdent As Long

 If bQuit Then Exit Function
 ' Open com port if not open
 If frmMain.msComm1.PortOpen = False Then frmMain.msComm1.PortOpen = True
```

```
  ' Clear input buffer
  frmMain.msComm1.InBufferCount = 0
  ' Send command
  frmMain.msComm1.Output = sCommand & Chr$(13)

  ' If command was to read channels then
  If sCommand = "#" Then
    ' Get data for each channel and store in appropriate place in AuxChannels array
    For iCounter = 1 To Len(WXSetup) - 2
      ' Which channel is being read
      iAuxChannel = Mid(WXSetup, iCounter + 2, 1)
      ' Get data
      nAuxChannels(iAuxChannel) = Val(CommInput)
    Next iCounter
    ReadMachine = ""
  ' Else some other command
  Else
    ReadMachine = CommInput
  End If

End Function

Function CommInput() As String

' Fetches returned data from the input buffer up until a carriage return is reached

Dim sLetter As String
Dim sBuffer As String
Dim nStart As Single
Dim nFinish As Single

  ' Set timeout variables
  nStart = Timer
  nFinish = nStart + 2

  Do
    DoEvents
    ' If data in input buffer then
    If frmMain.msComm1.InBufferCount > 0 Then
      ' Get character from input buffer
      sLetter = frmMain.msComm1.Input

      ' If character is a carriage return then response has been received
      If sLetter = vbCr Then
        Exit Do
      Else
        ' Else append character to sBuffer
        sBuffer = sBuffer & sLetter
      End If
    End If
    nStart = nStart + 1
  ' Loop until timeout or response has been found
  Loop Until Timer >= nFinish Or bExit

  CommInput = sBuffer

End Function
```

ModModule2

```
Option Explicit

Public Const HByteMask As Long = 255

Public Declare Sub Anjan Lib "vbio.dll" ()
Public Declare Sub Out Lib "vbio.dll" (ByVal port&, ByVal byt%)
Public Declare Function Inp Lib "vbio.dll" (ByVal portaddr&) As Integer

'8255 chip #1
Public Const IOPORTA As Integer = 432 '0x01B0
Public Const IOPORTB As Integer = IOPORTA + 1
Public Const IOPORTC As Integer = IOPORTA + 2
Public Const IOCONTROLPORT As Integer = IOPORTA + 3

'8255 chip #2
Public Const IOPORTA2 As Integer = IOPORTA + 4
Public Const IOPORTB2 As Integer = IOPORTA + 5
Public Const IOPORTC2 As Integer = IOPORTA + 6
Public Const IOCONTROLPORT2 As Integer = IOPORTA + 7

'8255 Specific Mode
Public Const MODE1AINBOUT As Byte = 188 '(10111100b) mode 1 with porta input, portb output

'Miscellanous constants
Public Const BIT0 As Byte = 1
Public Const BIT1 As Byte = 2
Public Const BIT2 As Byte = 4
Public Const BIT3 As Byte = 8
Public Const BIT4 As Byte = 16
Public Const BIT5 As Byte = 32
Public Const BIT6 As Byte = 64
Public Const BIT7 As Byte = 128
```

LIST OF PUBLICATIONS

- Δ Combined cardiac and respiratory monitoring using the Fibre Optic Respiratory Plethysmograph (FORP). Maletras F-X, Augousti A.T., Mason J. 2000 Proc. of Optics and Optoelectronics 2000, Loughborough, UK. Page 303

- Δ Construction and calibration of a new design of fiber optic respiratory plethysmograph (FORP). Maletras F.-X., Augousti A.T., Mason J. 2001 SPIE 4444 Proc. of Optomechanical Design and Engineering. San Diego, Cal. USA. Page 305.

- Δ Signal processing considerations in the use of the fibre optic respiratory plethysmograph (FORP) for cardiac monitoring. Maletras F-X, Augousti A.T., Mason J. 2001 Proc. of Sensors and Their Applications XI. London, UK. Page 312.