SEMEOTICONS

SEMEiotic Oriented Technology for Individual's CardiOmetabolic risk self-assessmeNt and Self-monitoring

---

D1.3 - Description of SEMEOTICONS reference dataset

Delivery Date	Month 18
Responsible	CNR
Work Package	1
Authors	CNR, FORTH, LIU, UCLAN, NTNU
Dissemination Level	PU

REVISION HISTORY

Revision no.	Date of Issue	Author(s)	Brief Description of Changes
0.0	May 11th 2015	Giuseppe Coppini	ToC circulation
0.5	June 18th 2015	All contributors	Collection of partial contributes
1.0	July 1st 2015	Giuseppe Coppini	Document integrated
1.0	July 6th 2015	Giuseppe Coppini	Document sent to IQR
1.1	July 8th 2015	Giuseppe Coppini	Minor changes according to quality review
1.1	July 8th 2015	Giuseppe Coppini	Quality approved by IQR
1.1	July 9th 2015	Giuseppe Coppini	Document sent to Project Coordinator
1.1	July 10th 2015	Project Coordinator	Document approved and sent to PO

List of Contributors

- CNR-IFC: Giuseppe Coppini, Aurora Morales, Riccardo Favilla, and Alessandro Mazzarisi.
- CNR-ISTI: Daniela Giorgi, Sara Colantonio
- FORTH: Franco Chiarugi,
- LIU: Tomas Strömberg, Marcus Larsson
- UCLAN: Bogdan Matuszewsky
- NTNU: Lise Randeberg

TABLE OF CONTENT

ABBREVIATIONS AND ACRONYMS

.avi	video/x-msvideo MIME Types
.bmp	Bit Map Picture MIME Types
.mts	Advanced Video Coding High Definition format MIME Types
.obj	Object (Wavefront) MIME Types
.wrl	VRML worlds 3D virtual reality object MIME Types
AGE	Advanced Glycosylated End products
AST	Aspartate AminoTransferase
AVCHD	Advanced Video Coding High Definition
BAI	Beck Anxiety Inventory
BDI-II	Beck Depression Inventory
BMI	Body Max Index
CM	Cardio Metabolic
CPU	Central Processing Unit
CVD	Cardiovascular diseases
DoW	Description of Work
DQP	Deliverable Quality Plan
ECC	Error Correcting Code
ECG	ElectroCardioGram
FINDRISC	Finnish Type 2 Diabetes Risk Score
FLI	Fatty Liver Index
GB	Giga Byte
GGT	Gamma-Glutamyl Transferase
HbA1c	Glycosylated Hemoglobin, Type A1C
HDL	High Density Lipoprotein
HF	High Frequency
HOMA	Homeostasis Model Assessment index
HR	Heart Rate
HRV	Heart Rate variability
HSCORE	Heart SCORE
HSI	Hyper-Spectral Imaging
HTTPS	Hypertext transfer protocol over secure socket layer
ISI	Insomnia Severity Index
JPEG	Joint Photographic Experts Group MIME Types
LASCA	Laser Speckle Contrast Analysis
LCTF	Liquid Crystal Tunable Filters
LDFDRS	Laser Doppler Flowmetry and Diffuse Reflectance method
LDL	Low Density Lipoprotein
LED	Light Emitting Diode
LF	Low Frequency
LOT-R	Life Orientation Test-Revised
MAC2-AF R	Motivation to change Physical Activity
MB	Mega Byte
Mpixels	Mega Pixels

MSI	Multi-Spectral Imaging
nm	NanoMeter
NUMi	Numeracy Understanding in Medicine Instrument
NVS	Newest Vital Sign
OEM	Original Equipment Manufacturer
pNN50	percentage of beats differing form the mean RR for more than 50 ms
PSS	Perceived Stress Scale
QR	Quality Reviewer
QRS	A combination of three waves (named Q, R, and S) and commonly seen on a typical electrocardiogram.
RAID	Redundant Array of Inexpensive Disks
RAM	Random Access Memory
RR	The time lapsing between two consecutive R-Waves in the ECG
SD	Standard Deviation
SDNN	Standard deviation of all normal RR intervals (those measured between consecutive sinus beats)
SF-12	Health Survey
SF12-MCS	Mental Health
SF12-PCS	Physical Health
SSL	Secure Sockets Layer
STEPS	WHO STEPwise approach to noncommunicable disease risk factor surveillance
TB	TeraByte
UV	Ultraviolet
WEB	Internet
WP	Work Package
XML	Extensible Markup Language

TABLE OF FIGURES

TABLE LEGENDS

EXECUTIVE SUMMARY

This report describes the data that were include in the SEMEOTICONS reference data set following the acquisition campaign held in Pisa in May 2014 (Task 1.3). The Campaign saw the participation of all partners involved both in semeiotic modelling of cardio-metabolic risk and in developing methods to extract computational descriptor of facial sign of CM risk. The work was started at Month 5 and terminated at month 18 when the protocol for the second acquisition campaign was defined. The Consortium effort in producing the data here described amount about to 19 PM.

The data are planned to serve for:
- developing and testing computational methods by: a) sample input data (video/images) as planned for the Wize Mirror implementation (D2.1.1), and b) reference ground truth made available by independent measurements (WP3,4 and 5),
- testing preliminary hypothesis during the elaboration of the final version of the semeiotic model of cardio-metabolic risk (WP1),
- driving the semantic integration in Task 6.1 and mainly T6.2 (WP6).

At the end of Task 1.3, a second acquisition campaign has been planned to extend the reference data. Similarly to the first acquisition campaign, it will be hold in Pisa in the last week of May 2015, will be participated by the same partners of 2014 campaign according to the same experimental protocol. To describe data of the second acquisition campaign we will release an addendum to this document by the end of September 2015.

PURPOSE AND SCOPE OF THIS DOCUMENT

This document describes the data SEMEOTICONS reference dataset collected during the acquisition campaign held in Pisa in May 2014 by the partners involved in WP1, WP3, WP4, WP5 and WP6 activities. The aim of this document is to produce a comprehensive description of the data, the way they were acquired and stored, as well as their retrieval. Relation between different data and project activities is also outlined. Data acquired from 23 volunteers, include a clinical and psychological characterization of subjects along with a set of multimedia data (video, images, signal, and 3D scans) scanned from volunteers' faces.

This document refers to all partners involved in modelling (i.e. partners working WP1 and WP6), methodological development (undergoing in WP3, 4, and 5) and technological validation (T8.6). Therefore, this report is eventually of interest for the entire Consortium.

The document is structured as follows:
- In the Introduction (page 11) we recall the aims of SEMEOTICONS reference dataset and summarize the 2014 acquisition campaign.
- In Section 2 (page 12) following a short recall of the experimental protocol we describe the general characteristic of the dataset population.
- In Section 3 (page 14) the clinical psychological characteristics of volunteers are reported.
- In Sections 4-8 we describe the acquisition setup and data format of multimedia data. In particular:
- Section 4 (page 20) deals with high frame-rate video with simultaneous ECG recording for HR and HRV estimation.
- Section 5 (page 24) describes the acquisition of still pictures of eye.
- Section 6 (page 25) reports the acquisition of 3D scans of the face.
- In Section 7 (page 26) describes the acquisition setup and data collected by Hyperspectral/Multispectral imaging for endothelial function assessment, skin cholesterol and skin AGE evaluation.
- In Section 8 (page 30) we describe the acquisition of images and movies during emotional stimulation.
- In Section 9 (page 31) the facilities for data storage and sharing are summarized.
- Finally, the Appendix I the Subject Information Form provided to volunteers at the enrolment phase is reported.

1   INTRODUCTION

SEMEOTICONS reference dataset includes a rich set of data gathered from volunteers and including, for each subject, i) a comprehensive medical characterization with CM risk factor assessment and ii) video, images, 3D facial scans, and signals reflecting the multi-sensing system to be implemented in the *Wize Mirr*or. The reference dataset is a key tool for SEMOTICONS activities. In fact, it provides medical data useful to test the hypothesis underlying the semeiotic model of cardio-metabolic risk (Task 1.1). Moreover, multimodal data are necessary for methodological development in WP3, 4, and 5 and are also expected to drive the design and implementation of the Virtual Individual Model in WP6. Finally, the gathered reference data has an impact on technological activity aimed to construct the *Wize Mirror*. In particular, the reference dataset has been planned to offer a common base for technological validation tasks (Task 8.6).

To collect the dataset, an acquisition campaign at the CNR campus of Pisa was planned, since the beginning of 2014. During the second half of April 2014 the medical researchers of the SEMEOTICONS Project started a screening on 26 healthy volunteers to be potentially enrolled in the acquisition campaign. In May, individual meetings with the candidates were started in order to explain the finalities of the Project and all the clinical/instrumental and biochemical analyses to be performed, in the greatest possible detail. Following the meetings, 23 of the 26 subjects accepted to be enrolled in the study. It is worth noting that subjects were selected so as to provide data in both normal and advanced CM risk states (e.g. hypercholesterolemia, diabetes, and hypertension not requiring drug treatment). After acceptance, they provided a written informed consent for their participation, an informed consent for the scientific use of the acquired images, as well as an informed consent for the publication of audio-video documents for dissemination purpose (The Information form provided to volunteers is reported in the Appendix I).

Data were acquired according to experimental protocol described in D1.4. Once enrolled, all subjects underwent blood sample analyses, including a metabolic risk factors testing. After the results of blood analyses were provided, all subjects underwent – as a first step - a comprehensive medical examination together with blood pressure measurement, assessment of oxygen saturation by pulse oximeter and body weight and height measurements. In the same session, or in a following one, the subjects were asked to fill all the questionnaires selected from current and validated literature by our psychologists and those by experts in nutrition. From May 19th to May 23rd, researchers of CNR, FORTH, UCLAN, LIU, and NTNU jointly participated to the acquisition campaign to gather multimedia data (images, videos, signals and 3D scans) from the 23 enrolled subjects. This phase of the acquisition protocol was designed and agreed by partners involved in WP3, 4, 5 and 6 in close connection with clinicians and psychologists. The first specification on Wize Mirror implementation (D2.1.1) was carefully taken into account when deciding the experimental setup. As to Hyperspectral/Multispectral imaging, the experimental protocol is also derived on the ground of preliminary investigations described in D1.2.

On average, each volunteer had to be available for data acquisition for 1 to 2 hours for 3 times. Face images/videos acquisition took up to 90 min. It is worth saying that, despite the complexity and the time consuming protocol, all subjects completed data collection and none of them complained for the type and duration of the different instrumental examinations performed during the campaign.

In this document, we describe the overall features of the volunteers (Section 2), their clinical and psychological characterization (Section 3), the multimedia data taken form face of each the subject (Sections 4-8). Finally, the facilities settled for sharing the data in the consortium are explained in Section 9.

2   DATASET POPULATION

To build the SEMEOTICONS reference dataset, the enrolled volunteers underwent the experimental protocol, which is fully described in D1.4 *Validation protocol*. For the sake of completeness we summarize the protocol in the next section.

1.1.   SUMMARY OF EXPERIMENTAL PROTOCOL

The criteria adopted in order to select the proposed volunteers were the following:

Inclusion criteria
- Healthy males and females in similar proportion
- Age   ≥25 and ≤60 years
- Willing to change their lifestyle by decreasing their body weight and/or increasing their physical fitness and/or improving their well-being following medical advice
- No overt disease
- Written informed consent

Exclusion criteria
- history of overt disease (including hypertension, diabetes, dyslipidemia requiring drug therapy)
- pregnancy or breast-feeding
- claustrophobia
- chronic medical treatment other than contraception

Each enrolled subject was clinically characterized according to the following items:

- Clinical Assessment (Section 3);
- Psychological, knowledge and nutritional tests (Section 3.3)
- Calculation of CM risk scores as a reference for individual well being (see Section 3.2).

In addition, during the joint phase of the acquisition campaign, the following multimedia data were obtained from each subject:

1. High frame-rate video with synchronous ECG (Section 4)
2. Eye images (Section 5)
3. 3D Face scans (Section 6)
4. MSI/HIS images (Section 0)
5. Images and movies during emotional testing (Section 8).

To perform multimedia data acquisition, the partners of the Consortium involved in the campaign designed a specific experimental setup.

1.2.   POPULATION DEMOGRAPHICS

The enrolled subjects included 16 (70%) male and 7 (30%) females. Age was between 25 and 61 years, mean age 45 years (SD 11 years). Female mean age was 47 years (SD 11 years), males mean age was 44 (SD 11 years). The age distribution is summarized in Figure 1
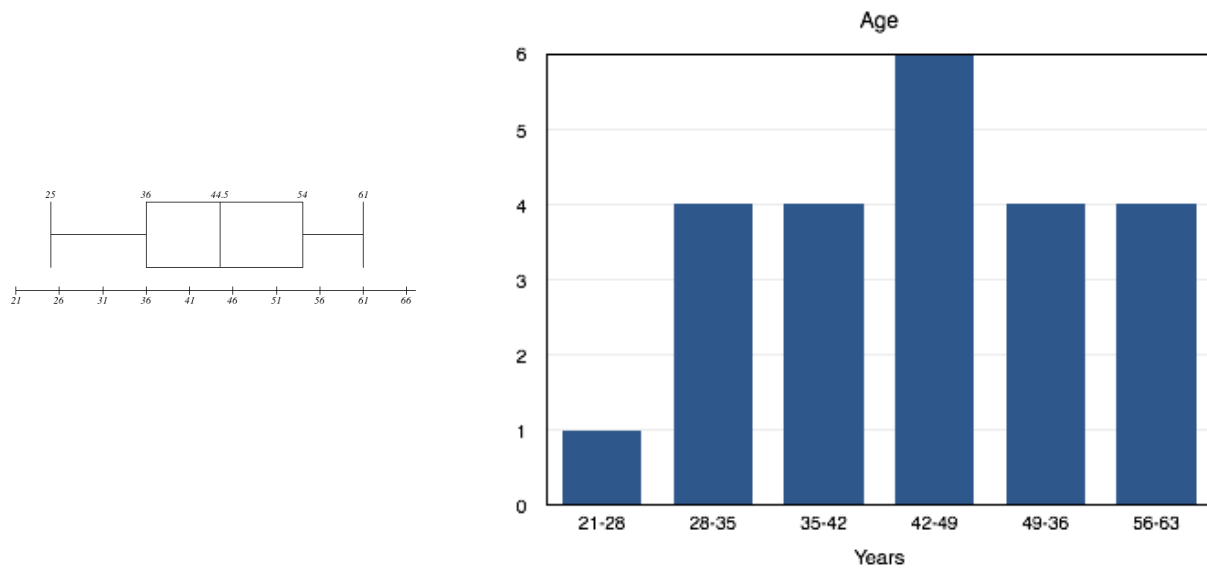
Figure 1. The left boxplot summarizes the age range of the enrolled subjects, while in the right panel the related histogram is drawn.

Additional general features of the Dataset population relates to education level, marital status, and occupation. In Figure 2, we summarize the major demographic features of the group of volunteers.
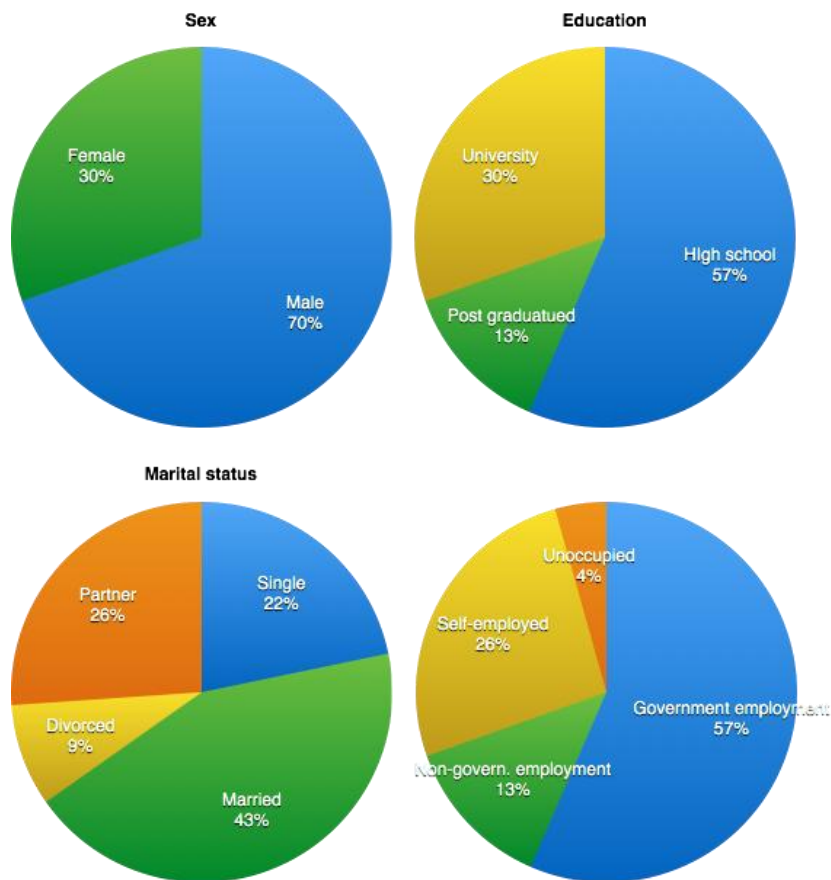
Figure 2. Main demographic parameters.

3   CLINICAL AND PSYCHOLOGICAL CHARACTERIZATION OF VOLUNTEERS

Clinical and psychological Characterization of enrolled subject is indented to provide reference values for the development of methods aiming to extract computational descriptors of facial signs of CM risk, that is most of the activities in WP3, 4, and 5. In addition, the clinical and psychological data provide a first test bed for the semeiotic model of CM risk in WP1 (Task 1.1).

3.1   CLINICAL ASSESSMENTS

For each volunteer, medical data about the following values were collected:

1. **Medical history**,
2. **Physical examination**: anthropometric measurements (height, weight, waist circumference, hip circumference) and vital signs (heart rate, respiratory rate, blood pressure)
3. **Body composition** (lean mass, fat mass) by BodPod (Cosmed)
4. **Resting energy expenditure** by indirect calorimetry by Quark (Cosmed)
5. Fasting **peripheral venous blood concentrations**: Total, HDL and LDL cholesterol, triglycerides, gamma glutamyltransferase (GGT), aspartate aminotransferase (AST), alanine aminotransferase (ALT), bilirubin, glucose, insulin, glycated haemoglobin (HbA1c), haemoglobin, creatinine
6. **Skin accumulation of Advanced Glycation End Products** (AGEs) by autofluorescence (AGE reader DiagnOptics)
7. **Endothelium-dependent vasodilatation** by peripheral arterial tonometry (ENDO-PAT, Itamar MEDICAL) and Laser Doppler flowmetry (Laser Speckle Contrast Analysis (LASCA, PERIMED)
8. **Heart rate and heart rate variability** by short term ECG monitoring
9. **Oxygen saturation** by pulse oxymetry
10. **Exhaled gas sampling**

3.2   RISK SCORE CALCULATION

Based on clinical characterization of volunteers, the following CM risk scores were computed for each volunteer:

- Heart SCORE (HSCORE)
- Fatty Liver Index (FLI)
- Finnish Type 2 Diabetes Risk Score (FINDRISC)
- Homeostasis Model Assessment (HOMA) index

3.3   PSYCHOLOGICAL, EDUCATION AND NUTRITION QUESTIONNAIRES

The following questionnaires were administered to volunteers to assess psychological, educational and nutritional aspects of an individual:

a) psychological well-being
- Life Orientation Test-Revised (LOT-R)

b) depression, anxiety, hostility
- Beck Anxiety Inventory (BAI)
- Beck Depression Inventory (BDI-II)

c) quality of life
- Perceived Stress Scale (PSS)
- SF-12 Health Survey
- Insomnia Severity Index (ISI)

d) lifestyle habits

- • WHO STEPwise approach to noncommunicable disease risk factor surveillance (STEPS)

e) motivation to change
- • Motivation to change Physical Activity (MAC2-AF R)
- • Motivation to change Nutrition (MAC2-AL R)

f) health literacy/numeracy
- • Newest Vital Sign (NVS)
- • Numeracy Understanding in Medicine Instrument (NUMi)

3.4 DESCRITIVE STATISTICS OF COLLECTED DATA

3.4.1 Clinical assessments

In the following table we summarize the central values (mean and median) and dispersion (SD) of the clinical data measured according to the protocol in D1.4.

Table 1. Mean, median values and standard deviation of the most relevant clinical features of dataset

Feature	Mean	Median	SD
Height cm	172.48	171.00	8.18
Weight Kg	83.30	75.00	31.27
BMI	27.57	25.93	8.29
Waist Circumf cm	97.91	92.00	22.09
HIP Circumf cm	105.30	101.00	19.77
Lean Mass % (imped balance)	73.92	74.80	7.04
Fat Mass % (Imped balance)	26.11	25.40	7.04
Lean Mass % BodPod	70.16	70.30	7.96
Fat Mass % BodPod	29.84	29.70	7.96
Body density Kg/l	1.03	1.03	0.02
Thoracic Gas Volume l	3.80	3.63	0.82
Energy Expenditure Kcal/day	1768.17	1696.00	377.34
Delta% of EE predicted	10.60	11.40	5.46
Respiratory Quotient	0.85	0.83	0.08
VE l/min	31.13	32.60	4.90
Oxygen Saturation %	97.70	98.00	0.88
Heart Rate (bpm)	64.74	67.00	9.86
Systolic Pressure mmHg	128.70	125.00	11.00
Diastolic Pressure mmHg	77.30	75.00	7.50
QT interval msec	383.91	390.00	25.94
SDNN msec	58.74	53.00	30.07
pNN50 %	8.18	4.25	9.93
HF nu x Hz	101.97	91.00	60.05
LF nu x Hz	158.98	152.45	66.83
LF/HF	2.00	1.90	1.21
RHI	2.23	2.16	0.60
lnRHI	0.77	0.77	0.26
AI %	8.00	10.00	14.92
AI@75 %	-0.09	-4.00	15.10

AGE_AF1	2.27	2.20	0.48
HB gr/dl	14.71	15.00	0.98
RBC ul	4.80E+06	4.95E+06	9.83E+05
WBC ul	5.91E+03	5.64E+03	1.34E+03
Platelets ul	2.50E+05	2.39E+05	8.73E+04
Bilirubine mg/dl	1.06	0.93	0.39
GOT mU/ml	29.48	23.00	29.15
GPT mU/ml	27.83	25.00	13.36
GGT U/L	29.00	29.00	10.90
Cholestorol mg/dl	232.00	224.00	51.22
HDL mg/dl	52.13	49.00	17.21
LDL mg/dl	155.83	152.00	46.41
Triglicerides mg/dl	123.87	86.00	88.46
Glucose mg/dl	94.13	91.00	8.46
HBa1c mmol/mol	37.70	38.00	2.53
INSULIN pmol/l	11.15	10.53	7.03
Creatinine mg/dl	0.95	0.97	0.14

Table 2 shows a summary of the presence of hypertension, hypercholesterolemia, the tobacco usage (smokers) and the fraction of overweight and obese subjects, according to BMI value. Values are given as frequency and percentage.

Table 2. Distribution of some major CM risk factor in the dataset population

Risk factor	Presence	Freq	%
Hypertension	No	20	86.96
	Yes	3	13.04
Hypercholesterolemia	No	10	43.48
	Yes	13	56.52
Tobacco use	No	16	69.57
	Yes	7	30.43
BMI	Normal weight	7	30.43
	Overweight	13	56.52
	Obesity	3	13.04

3.4.2 Risk score calculation

Following data collection, for each subject the values of HSCORE, HOMA index, FLI and FINDIRISK were computed, while central values (mean, median) and dispersion (SD) of these scores are shown in Table 3.

Table 3. Cardio-metabolic risk score: mean, median and standard deviation

Score	Mean	Median	SD
HSCORE	1.06	0.54	1.35
HOMA Index	2.57	2.55	1.54
FLI	44.93	46.86	30.02
FINDRISC	7.04	7	3.86

Plotting histograms of the related distributions allow a better insight of their meaning in SEMEOTICONS population. As it can be expected for a set of prevalently young and overall healthy subjects, HSCORE that estimates the probability of a major cardiovascular event in the following ten years is concentrated in a range near zero, denoting very low probability of such events for the population at hand. Differently, HOMA Index, FLI and FINDIRSC exhibit a broader distribution. The latter scores are focused to measure metabolic risk, though with different perspectives: HOM Index should describe insulin resistance, FLI was conceived as an indicator of liver fat, and FINDIRISC is the score (specifically conceived for Northern Europe populations) for assessing the risk of diabetes.
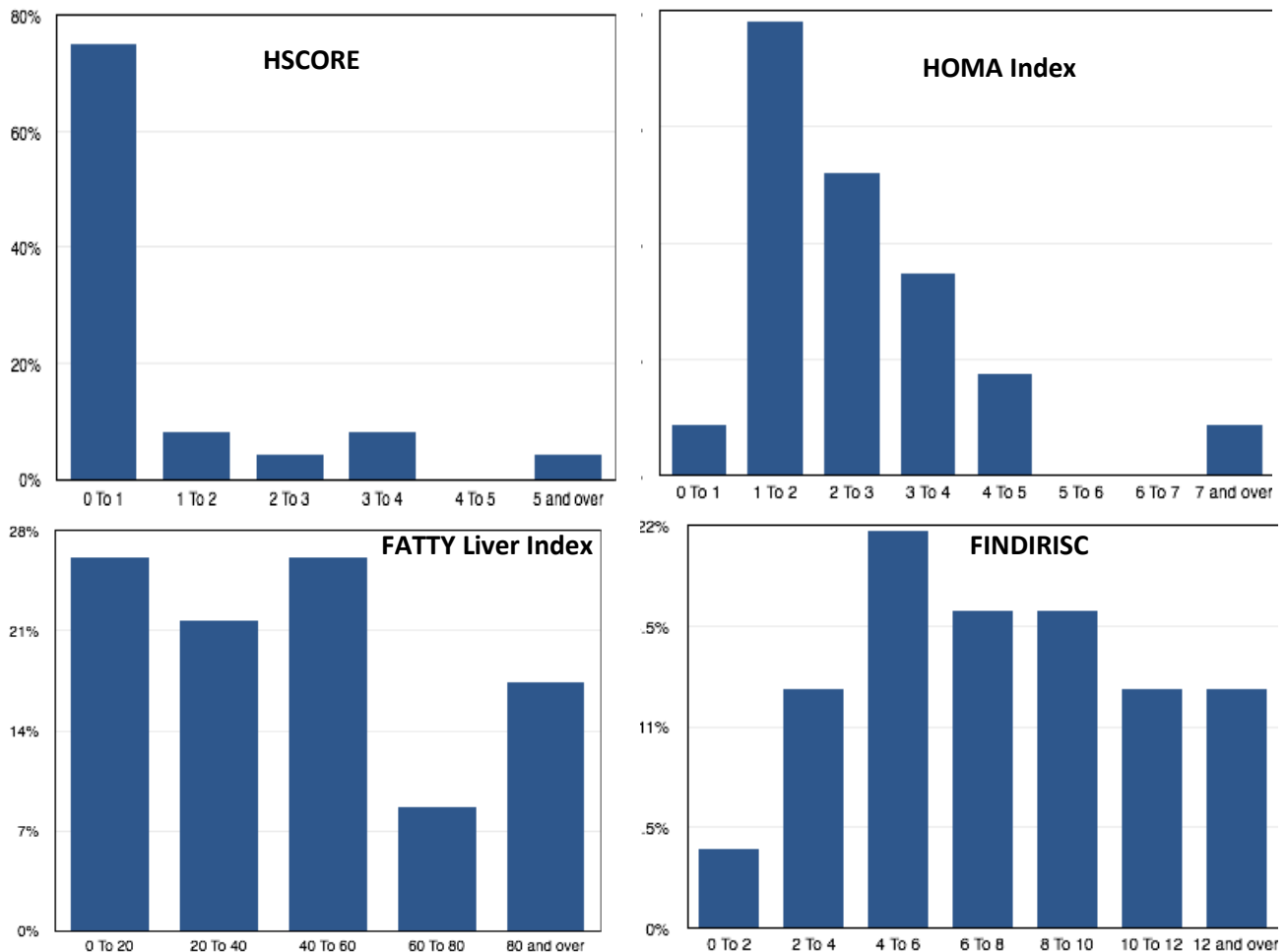

Figure 3. Distribution of risk score in the reference dataset population

17

3.4.3 Psychological, education and nutrition questionnaires

A huge quantity of data was collected from volunteers by means of validated questionnaires (see *D1.4 Clinical Validation protocol*). In Table 4, we report a summary description of data set values for the set of behavioural and psychological features, including lifestyle habits (WHO Steps questionnaire), anxiety and depression (BAI and BDI-II), insomnia (ISI) health literacy and numeracy (NVS and NUMI respectively), as well as motivation to change (MAC2)

Table 4. Behavioral and psychological features of SEMEOTICONS subjects

Feature		Freq	%	Cum %
Tobacco use	Non-smoker	15	65.22	65.22
	Smoker but not every day	4	17.39	82.61
	Every day smoker	4	17.39	100
Alcohol use	Non-drinker	3	13.04	13.04
	Drinking in the last year	6	26.09	39.13
	Drinking in the last month	14	60.87	100
Nutrition	No risk	2	8.7	8.7
	Medium risk	11	47.83	56.52
	High risk	10	43.48	100
Physical activity	Low level	14	60.87	60.87
	Medium level	9	39.13	100
Anxiety	No anxiety	20	86.96	86.96
	Moderate anxiety	1	4.35	91.3
	Severe/moderate anxiety	2	8.7	100
Depression	No depression	20	86.96	86.96
	Mild depression	2	8.7	95.65
	Severe depression	1	4.35	100
Stress	No stress	13	56.52	56.52
	Stress	10	43.48	100
Insomnia	No insomnia	19	82.61	82.61
	Sub threshold insomnia	3	13.04	95.65
	Moderate severity clinic insomnia	1	4.35	100
Health literacy	Medium level	4	17.39	17.39
	High level	19	82.61	100
Health numeracy	Low level	2	8.7	8.7
	Medium level	12	52.17	60.87
	High level	9	39.13	100
Motivation to change - physical activity	Contemplation	9	39.13	39.13
	Determination	4	17.39	56.52
	Action	2	8.7	65.22
	Maintenance	8	34.78	100
Motivation to change - nutrition	Pre-contemplation	2	8.7	8.7
	Contemplation	7	30.43	39.13
	Determination	1	4.35	43.48
	Action	2	8.7	52.17
	Maintenance	11	47.83	100

In Table 5 we report the central values (mean and median) and dispersion measurements (SD) of perceived health both physical and mental (SF12 – PCS and MCS respectively), perceived stress (PSS) and optimism. These scores are intended to describe the self-perceived wellness of an individual. As such, they are an

index of the wellness and are expected to provide useful hints for the computation of SEMOTICONS wellness index.

Table 5. Score about perceived wellness: SF12 scores for physical and mental health, Perceived Stress and LOT-R score

Score	Mean	Median	SD
Physical health (SF12 - PCS)	51.35	55.20	6.95
Mental health (SF12 - MCS)	44.93	46.86	30.02
Perceived Stress Scale	11.74	12.00	5.90
Optimism (LOT R)	18.65	18.00	2.40

4 HIGH FRAME-RATE VIDEO WITH SYNCHRONOUS ECG

These data were acquired in order to have a a set of high frame rate videos (see deliverable D2.1.1 *Initial specification of system requirement and functionalities*) coupled with synchronous ECG, to be used as a gold standard for heart rate (HR) and Heart Rate Variability (HRV). These data impact directly on T4.4 and are relevant for methods devoted to emotional characterization under development in WP5.

4.1 ACQUISITION SETUP

The acquisition setup (visible in Figure 4) included:

- An *ad hoc* ECG device assembled at CNR-IFC by using an EG05000 OEM board from Medlab [[1]].
- A uEye Gigabit Ethernet Camera from IDS with an e2V CMOS sensor having 1.31 MPixel, a maximum frame rate 50.0 fps with 1280 x 1024 sensing matrix and 10 bits per pixel [[2]]. The camera is equipped with an optical band-pass filter, with central wavelength at 500 nm (±20 nm) placed ahead the camera lens.
- An approximately uniform lighting source built with strips of white LEDs mounted on a wood frame.
- Application software developed at CNR-IFC for synchronized data acquisition, processing and storage of high frame rate videos and ECG signal.
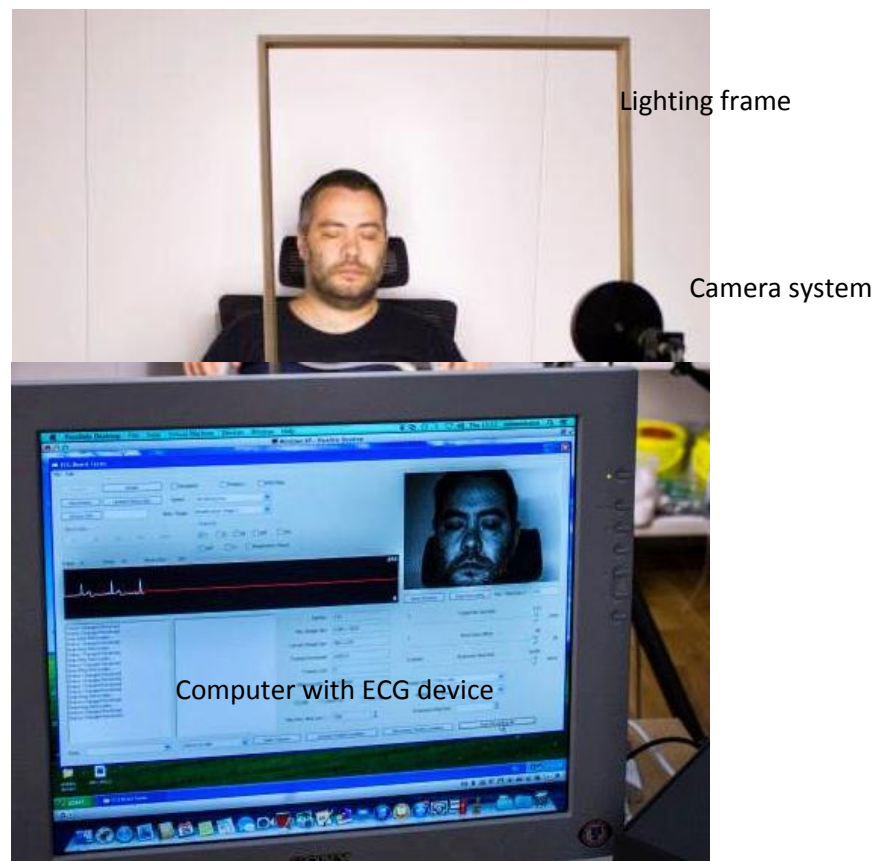

Figure 4. Acquisition setup for ECG synchronous High frame-rate

4.1.1 Application Software

In order to manage the hardware devices and provide a synchronous recording of ECG and video signal, custom application has been designed and implemented. A screenshot of the window dedicated to the configuration of devices, previewing and registration of both ECG and video signals.
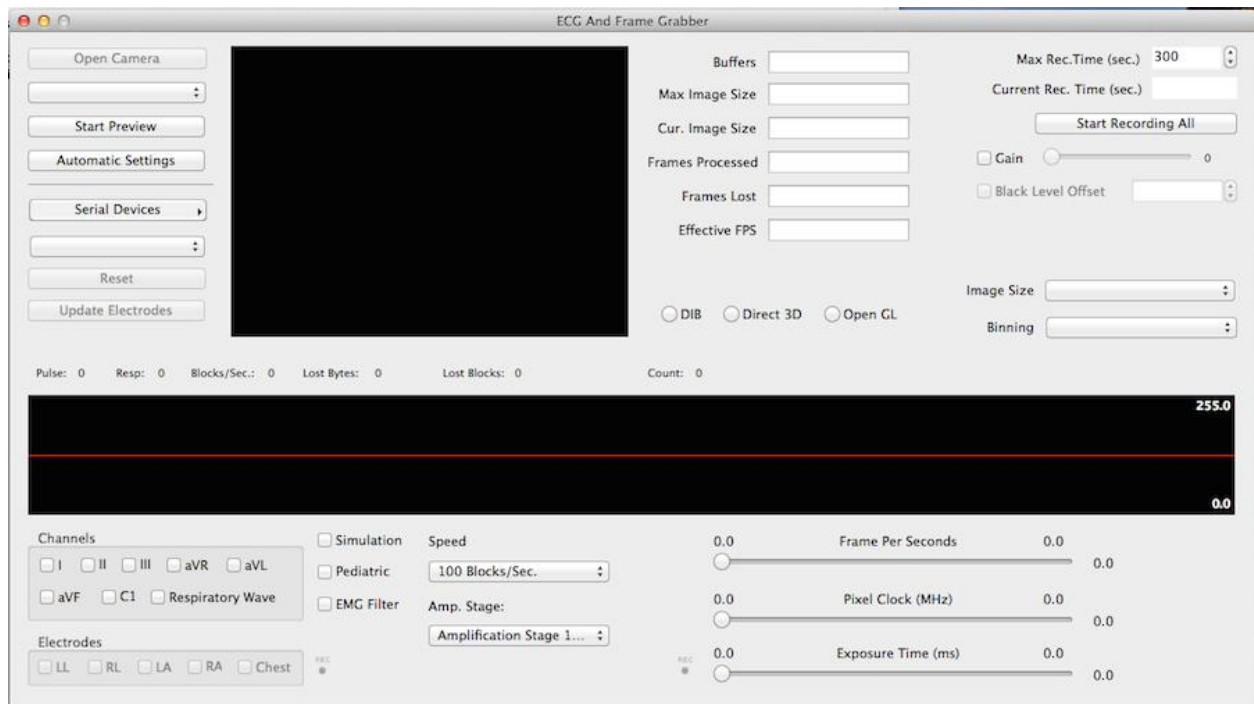
Figure 5.Screenshot of the application software acquisition interface

As showed, the software allows selecting the appropriate devices, configure them according to their hardware features, while real-time previewing changes were performed to the configuration in use.

4.1.2   Acquisition Protocol

For each of the 23 subjects in the dataset population, we applied the following experimental protocol:

- ECG device was setup to transmit 300 waveform blocks/second, with only the first channel enabled, corresponding to ECG lead I. Furthermore, the amplification of the waveforms was set to a value of 3.
- the video camera was connected directly to a Mac mini PC (Intel Core i7 quad-core 2.4 Mhz, 16GB) via a Gigabit Ethernet port; the size of recorded images was set to 704 x 448 pixels with a two-pixel binning and a depth of 8 bits per pixel. With these camera settings, we achieved a rate of 133 frames per second. The camera was positioned on an tripod in front of the subject in such a way that only his/her face is imaged (camera-subject distance was 1 meter, the camera mounted a TUSS Mega-pixel Fixed Focal Lens 2/3 16mm F1.4. The video signal was stored uncompressed to disk.
- the lightening frame was placed in front of the subject at a distance of about 60 cm.
- the duration of the recording was fixed to 5 minutes and the subject was advised to stay as still as possible for the entire duration of the recording

4.1.3   Description of data

Both ECG and video data collected during each recording session have been stored on disk in a folder containing the following files:
- a file named *ueyeseq.raw* storing the raw uncompressed video data. The presence of this file is mandatory and enables the application to recognize a folder as containing valid data. No additional information is stored into this file: the characteristics of the recorded video stream are stored into the *xml* file described later
- a file named *ecg.raw* (optional) that, like name suggests, stores ECG data. This file contains both the ECG waveform blocks (it contains the marker 0xF8, a byte with the number of recorder lead and

one byte per lead) as registered during the acquisition session and the pulse block (one byte preceded by the marker 0xFA) that is transmitted at each detected pulse (QRS complex).

– an xml file named acqInf.xml providing the acquisition parameters of both video and ECG data characteristics allowing to load and playback data. These include the ECG sampling rate, the ECG leads recorded, the camera frame rate, the image size and the pixel depth.

All of the data stored during the acquisition session can be loaded by application and used to evaluate Heart Rate and Heart Rate Variability values from the analysis of the video signal (see Figure 6)

Figure 6. Visualization of acquired data.

In the upper part of the window, the registered video is displayed, along with general information about the data; it is possible to analyze a part of the image by specifying a region of interest (the red rectangle in the figure) or the entire image. In the middle part of the window, the graph of the registered ECG data is displayed, with the vertical green lines, indicating the time at which a pulse has been detected. In the lower part of the window, the graph of the grey level variations over time is displayed; such a graph has been obtained in every image by mediating the values of each pixel within the region of interest (or the entire image). In order to regularize the signal, minimization of the motion artifacts and/or changes in illumination conditions were performed, since video data can be filtered, which ease peak detection and helps reducing the number of spurious peaks.

In Table 6 we report some relevant parameters computed from the ECG reference signal, describing HR and HRV for the subjects of the Acquisition Campaign. The table provides the mean RR, the related standard deviation (SDNN), the value of pNN50 indicating the percentage of beats differing from the mean RR for more than 50 ms, the low frequency power (LF), the high frequency power (HF) the ratio LF/HF, and the ECG derived respiration rate. The overall dataset acquired can be suitably employed in order to test methods for assessing HR and HRV from videos.

Table 6. Data obtained from ECG signal fro HR and HRV descriptors in the reference dataset.

Subject ID	Mean RR (ms)	SDNN (ms)	pNN50 (%)	LF (nu)	HF (nu)	LF/HF	EDR (Hz)
1	1013.10	41.56	3.63	82.91	17.04	4.87	0.15
2	939.40	94.26	19.75	86.62	13.36	6.48	0.19
3	905.54	83.56	32.52	66.11	33.84	1.95	0.21
4	997.90	39.94	21.28	35.37	64.61	0.55	0.23
5	901.13	53.49	30.89	48.84	51.09	0.96	0.16
6	1013.50	99.93	21.84	58.07	41.90	1.39	0.18
7	919.15	40.06	10.94	57.26	42.59	1.34	0.22
8	1111.10	53.06	36.36	20.54	79.43	0.26	0.21
9	854.83	25.17	3.21	51.47	48.20	1.07	0.22
10	834.38	26.92	0.00	58.16	41.79	1.39	0.22
11	740.60	33.63	0.00	75.28	24.68	3.05	0.27
12	874.42	110.76	63.39	40.72	59.24	0.69	0.24
13	775.84	34.08	0.26	82.11	17.87	4.60	0.25
14	786.24	27.49	0.00	88.04	11.94	7.37	0.33
15	1090.70	51.27	52.21	24.71	75.25	0.33	0.25
16	961.49	47.02	6.84	73.71	26.22	2.81	0.28
17	916.26	42.51	7.14	55.61	44.25	1.26	0.23
18	711.10	32.98	1.66	71.41	27.32	2.61	0.39
19	745.64	43.84	4.55	62.48	37.50	1.67	0.16
20	1131.70	52.63	53.99	43.17	56.83	0.76	0.26
21	823.09	25.87	9.29	25.50	74.28	0.34	0.24
22	1099.90	56.82	20.57	44.08	55.91	0.79	0.21
23	885.88	45.24	1.81	86.56	13.42	6.45	0.30

5   EYE IMAGES

The eye images were collected to support Task 3.4 "Iris image analysis for indicative detection of abnormal level of cholesterol". The goal of that task is to test the hypothesis that the information contained in an image of the eye, and of the iris in particular, could be linked with clinical risk scores. This requires developing iris image processing techniques, including iris segmentation and normalization. The processing should facilitate further research on the analysis of patterns present in iris images (e.g., arcus cornealis) and their link to corneal lipid accumulation.

5.1   ACQUISITION SETUP

Two cameras were used:
– a CANON 5D, with lens CANON EF 24-105, and
– a CANON EFS, with lens CANON EFS 18-55.

The acquisition set-up included a photography umbrella with white lighting. A professional photographer supported the study of the best set-up conditions.
As the first experiments revealed the presence of artefacts in the images due to the lights (i.e., a white dot in the iris area), we also experimented with an in-house made device, derived from a plastic glass, placed around the eye area.

5.2   DESCRIPTION OF DATA

For each of the 23 volunteers, 10 images were saved, 5 of which acquired with the CANON 5D, and the other 5 with the CANON EFS.
Each group of 5 images was made up of a picture of the whole face, 2 pictures of the left eye and 2 pictures of the right eye, as the eye pictures were acquired with and without the in-house made device.
The images were saved in both raw and JPEG format. Sample images are shown in Figure 7.
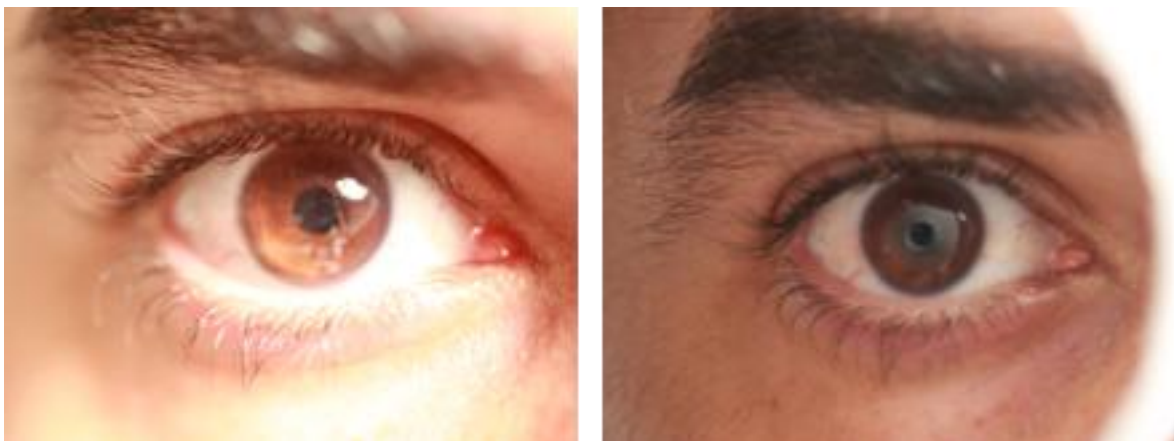

Figure 7. Image of the right eye of a subject. Left: direct exposure, right: plastic shielding around the eye.

6   3D FACE SCANS

The acquisition of 3D faces relates to WP4 "3D models construction and characterization". The 3D reconstructions are used for bio-morphometric and multispectral data analysis (Task 4.3). In conjunction with the data acquired during the second acquisition campaign, the data also facilitate evaluation of the inexpensive 3D scanner developed specifically for the mirror (see also deliverable D4.1.1 *3D geometric reconstruction subsystem)*.

6.1   ACQUISITION SET-UP

Since the acquisition campaign took place at a very early stage of the project, when the 3D reconstruction system was not available as a prototype yet, we used a commercial scanner.
The scanner was an inSpeck 3D Mega Capturor II, with field of view 568x455mm; lateral resolution of up to 0.44mm; depth resolution of up to 0.69mm; acquisition time 0.7sec.

6.2   DESCRIPTION OF DATA

The data stored include the raw data produced by the scanner, the 3D data converted in standard formats (.obj and .wrl), and the .bmp images of the acquired faces. For each face, there are three 3D meshes, corresponding to the left half of the face, the right half, and the front part, respectively. The average number of mesh vertices is ~30,000, and the average number of mesh polygons is ~60,000.
Figure 8 shows example meshes on a dataset subject.

Figure 8. Sample meshes generated for a subject of the reference dataset.

7 MULTISPECTRAL / HYPERSPECTRAL IMAGING

Acquisition of Multispectral/Hyperspectral images has been planned to provide data from developing and testing methods for a) the assessment of endothelial function with thermal stimulation (Task 3.1), b) the evaluation of skin cholesterol deposition (Task 3.2), and c) the evaluation of skin AGE by auto fluorescence (Task 3.3).

7.1 ACQUISITION SETUP, MSI AND HSI HARDWARE FOR ENDOTHELIAL FUNCTION, AGE ACCUMULATION AND CHOLESTEROL LEVEL

During the Pisa data acquisition campaign, in May 2014, a twin camera MSI system (wavelength bands: 355nm and 475nm) was used for remotely quantifying the AGE product concentration on the skin (Figure 9). One set of computer controllable white LEDs and one set of computer controllable UV LEDs were used.
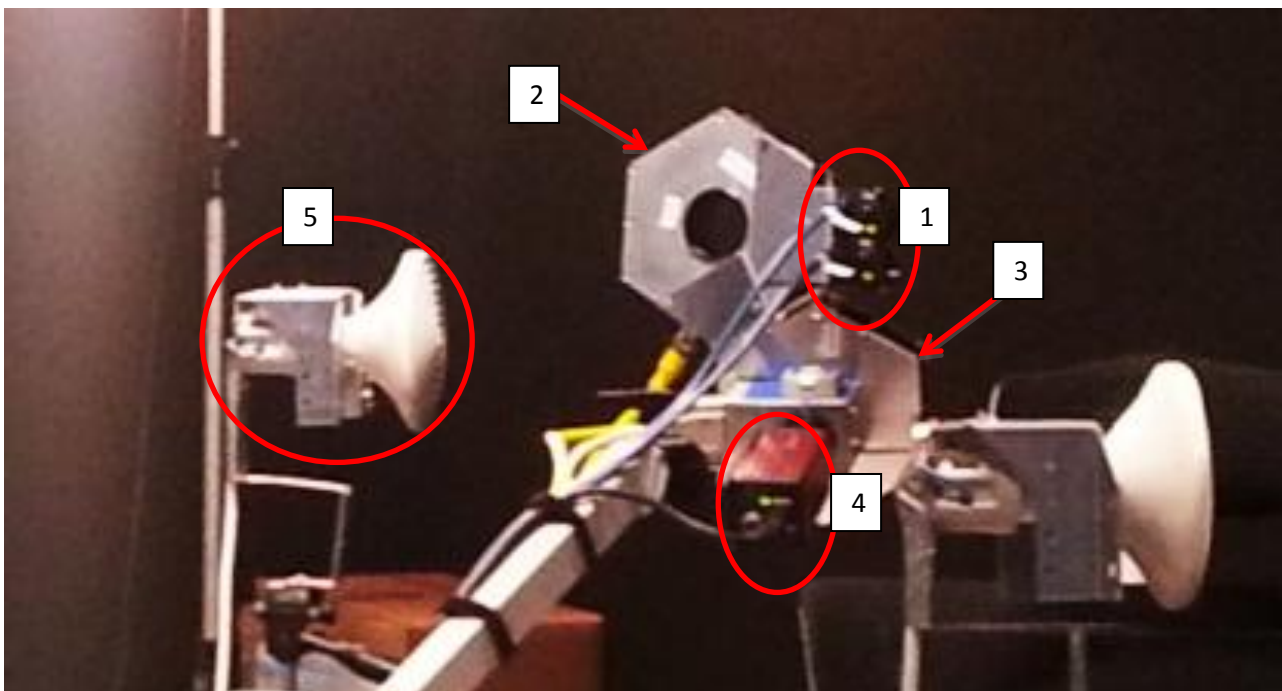

Figure 9. MSI setup with two cameras (1), white (2)- and UV-LED (3)-light sources, hyperspectral imaging camera (4) and ceramic IR-heaters (5).

The cameras are USB3.0 FL3-U3-32S2M-CS cameras from Point Grey, each equipped with a 8mm/f1.4 Tamron M118FM08 lens. These cameras are capable of capturing images with a maximum resolution of 3.2MP. With on-chip pixel binning capabilities, this resolution can be decreased to 0.8MP for an increased signal-to-noise ratio and decreased demand for data transfer and image processing capabilities. These cameras are also equipped with hardware and software syncing capabilities. The MSI cameras are equipped with high quality optical filters that captures selected wavelength band. One fluorescence band pass filter at 475nm (+- 25nm, 6 OD, ø25mm; Edmund Optics, Stock No. #86-353), one band pass filter at 355nm (53nm FWHM, ø25.4mm; Edmund Optics, Stock No. #46-048, 64€). The MSI system was equipped with two sets of controllable LED lamps: A white LED (Smart Vision Light Inc, L300-WHI-W), a 365nm UV LED (Smart Vision Light Inc, L300-365-W).

The HSI system consisted of a camera equipped with a tunable LCTF band pass filter, optical lenses and a white LED ring light. The illumination and the image capturing was synced and controlled from a computer allowing for an automated capturing of both white light images and dark images for wavelengths in the range 400-720nm. Only spectral data in the 420-720nm could be considered reliable as the luminescence of the white LED and the sensitivity of the HSI camera was too low for shorter wavelengths.

7.2   ASSESSMENT OF ENDOTHELIAL FUNCTION

7.2.1 Reference method and pointwise and imaging methods for measuring microcirculation blood flow

Endothelium-function can be assessed in clinical practice using e.g. peripheral artery tonometry or by studying the microcirculatory blood flow after local heating.

As reference method for endothelial function, peripheral arterial tonometry (PAT) was used (Figure 10).
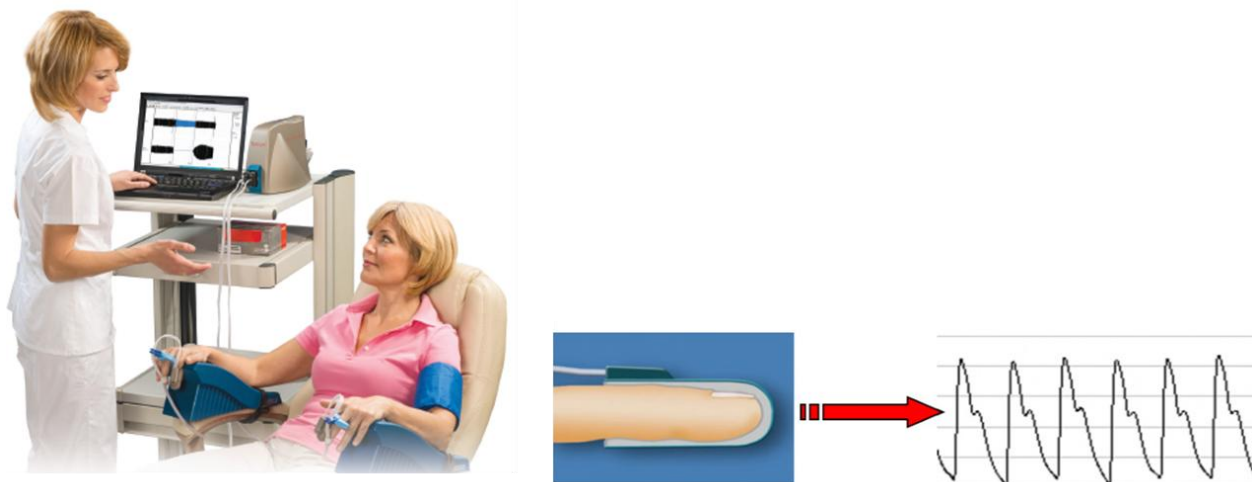

Figure 10. Peripheral arterial tonometry measuring finger arterial pulse volume during post-occlusive reactive hyperemia (PORH). Itamar Medical [[3]].

The digital/finger hyperaemia response after 5 min of systolic occlusion was measured as the reactive hyperaemia index (RHI; normal > 1.67; see also [[3]]).

A fiberoptic method was used for point-wise measurement of skin microcirculation during local heating (Figure 11). The probe and a small local skin heater were positioned at the volar forearm skin in an area free from visible veins. The probe is held in position using double adhesive tape. The integrated Laser Doppler Flowmetry and Diffuse Reflectance method (LDFDRS), is described in Strömberg et al 2014 [[5]] and the data analysis in Fredriksson et al 2013 [[6]].

Figure 11. Fiberoptic probe with local heater for measuring skin microcirculation response Perimed AB, Järfälla, SWEDEN [[4]].

Microcirculation blood flow imaging was done using LASCA (LAser Speckle Contrast Analysis). With the LASCA instrument (PeriCam PSI, Perimed AB, Sweden [[4]]) it is possible to study changes in the microcirculation perfusion by analyzing Doppler related effects in light that has been backscattered from a laser illuminated tissue. The LASCA perfusion measure is therefore mainly sensitive to changes in the blood

27

flow speed. Even though this technique is based on the same Doppler principle as the LDF technique described above, one cannot directly compare these measures one-to-one.

7.2.2   Protocol

The protocol involved at least 10 min of acclimatization while sitting in a comfortable chair in a room that had a temperature of at least 23°C (mean = 25.5^{o}C). The fiberoptic probe was non-heated for at least 5 min (baseline recordings), and then forearm skin was heated to 44°C during 15 min. During this period the subjects' facial skin was heated for up to 15 min using ceramic heaters at about 40 cm distance, aiming at a facial skin temperature about 40°C. HSI images were recorded in baseline and at 5 min interval during heating. LASCA perfusion was recorded for about 30 seconds prior each of the HSI image, from baseline until after 15 of facial heating.

The main motive for this procedure was to:
1. Evaluate the remote heating technique and the heating protocol.
2. Design MSI algorithms for estimating blood saturation and concentration using a subset of the HSI data.
3. Compare the HSI/MSI and LASCA measures.
4. Compare the measured microcirculatory response to other clinical reference data for endothelium function evaluation.

7.2.3   Description of data

The fiberoptic probe recorded LDF and DRS spectra as described in Strömberg et al 2014 [[5]]. Spectral data were recorded at 1 or 5 Hz, for the protocol of 25 min using the LIU/Perimed software DemonstratorWindows. Data files are up to 7 MB.

LASCA recordings of 2 min, recording 2 images/ s, give files of up to 40 MB.

HSI images of 32 wavelenghts and 1.4 Mpixel at baseline and after heating, give files of about 110 MB each.

7.3   ASSESSMENT OF SKIN AGES

7.3.1   Reference method and HSI for AGE assessment

As reference method, the AGE reader from Diagnoptics Technologies B.V., The Netherlands [[7]], was used (Figure 12).

Figure 12. AGE reader by Diagnoptics Technologies B.V., The Netherlands [[7]].

The HSI system had too low sensitivity below 420 nm. Hence, no direct imaging of the UV diffuse reflectance can be achieved with this setup. Therefore, no normalisation with the excitation luminance can be achieved using this data, which eliminates the possibility of estimating the AGE concentration using the HSI setup. The HSI data was, therefore, used for quantifying the spectral appearance of the AGE auto fluorescence in order to determine the feasibility of using a UV-MSI setup for measuring the concentration of AGE products in skin.

7.3.2   Protocol

A baseline image was recorded using UV-illumination by the MSI and HSI systems.

28

7.3.3 Description of data

An MSI image series of 10 dark images and 10 UV images, each sized 0.8 Mpixel, gives a total file size of about 25 MB. HSI images of 32 wavelengths and 1.4 pixels give files of 110 MB.

7.4 ASSESSMENT OF SKIN CHOLESTEROL

7.4.1 Protocol

HSI data were recorded during white light illumination.

7.4.2 Description of data

HSI images of 32 wavelengths and 1.4 Mpixels during white light illumination, give files of 110 MB.

8   IMAGES AND MOVIES DURING EMOTIONAL TESTING

8.1   ACQUISITION SETUP

The experimental procedure is aimed to record videos, conveying meaningful information to detect anxiety/stress/fatigue states from facial video recordings.

The participants were seated in front of a computer monitor while the cameras were placed at a distance of about 50 cm. At the beginning of the procedure, the participants were informed about the different meaning of the terms anxiety, stress, and fatigue.

The equipment included 2 color video cameras: one with high resolution and low frame rate, provided by CNR, and one with low resolution and high frame rate, provided by FORTH. An environment with high diffuse light was set for all the experiments.

8.2   DESCRIPTION OF DATA

For each of the 23 participants, 13 videos were recorded. The videos had durations of 30 sec, 1 minute and 2 minutes with a resolution of 526x696 pixels at 50 fps and with resolution of 1920 x 1080 pixels at 50fps. The format of the former videos was uncompressed Audio Video Interleaved (.avi file) and the format of the latter was AVCHD (Advanced Video Coding High Definition) (.mts file).

In some of the video recordings, the volunteers were asked to simulate a state of stress/anxiety/fatigue. In other video recordings, visual stimuli were provided to participants to elicit the emotions. The stimuli included stressful stimuli consisting of non-calming images presentation, stressful videos presentation, and performance of difficult mental tasks. The stimuli were moderate, according to the guidelines of the study's ethical committee. After watching each video/image, each participant filled a self-report questionnaire reporting the feeling experienced during the video/image. This was done by using a rating from 1 to 5, where 1 stands for "Relaxed" and 5 for "Stress or Anxiety".

The overall procedure is depicted in Table 7.

Table 7. Procedure and description of data for emotional testing

A/A	Video length (min)	State
1	0.5	Reference period
2	1	Anxiety/Stress
3	1	Reading text
4	0.5	Neutral
5	0.5	Anxiety
6	0.5	Stress
7	0.5	Fatigue
8	1+1	Relaxed + Stress viewing images
9	2	Mental task performance
10	0.5	Neutral
11	1.5	Relaxed
12	2	Anxiety/Stress
13	2	Anxiety/Stress

9   DATA STORAGE AND SHARING

Two different repositories were prepared for data storage and sharing: the first one for clinical and psychological data, the second one for multimedia data.

9.1   CLINICAL AND PSYCHOLOGICAL DATA

For the purpose of collecting, storing, consulting and sharing data between partners, an electronic record was built that comprises of 320 fields. This database is shared on a server hosted at CNR-IFC in Pisa through a public network. Access to data is performed via SSH authentication through a conventional WEB browser, thus each partner is provided with a personal account with username and password.

A relational database has been developed, that includes both medical baseline evaluation and psychological evaluation. Data were arranged in sections in order to simplify interpretation. Each user may access the collected data through the main page, which allows patient selection. Once selected, it is possible to access to all the sections relative to the same patients; otherwise, it is possible to navigate between patients for a specific section. Sample screenshot of the database interface are shown in Figure 13.

Figure 13. Screenshots taken from the database interface.

Researchers involved in the SEMEOTICONS reference database protocol are allowed to access the collected data in anonymized form.

The server used to store anonymized data is deployed in the CNR Physiology Clinical Institute Server Farm facility room. The database used is "FileMaker 12 Server, FileMaker, Inc". To safely treat data, the server uses a journaled MacOS extended protected file system. Passwords and data transmission, across local area network and across the Internet, are exchanged according to SSL protocol. Automatic data backup will be done from the server scheduler, storing a full fresh copy each hour, day and month according to round robin algorithm. Authorized user can export data in the most common tabular formats suitable for statistical processing.

9.2   MULTIMEDIA DATA

To guarantee secure data storage and efficient data access, the data acquired during the acquisition campaign were saved on a server with 128 GB of ECC RAM and eight disks of 4 GB each, managed by an

Adaptec Series 6 - 6G SAS/PCIe 2. The server is based on a dual-core Intel(R) Xeon(R) CPU E5-2670 v2 that allows the execution of 40 parallel tasks. It is physically located in Pisa, at the CNR premises.

The controller manages the disks using a RAID 6 configuration that gives 24 TB available and a resilience of two disks: in other words, there is a data loss only after the third disk is broken. This solution guarantees a low data loss risks considering the RAID is plugged to two independent uninterruptible power supply. RAID technology also provides data redundancy: the parity is distributed across all the drives in an array but it does not take up as much space as a complete mirror.

The server allowed us to handle the 2 TB of data acquired during the campaign. The data were organized both *per subject* and *per data*, so that one can easily access either all the data of a single subject or a particular type of data (e.g., 3D reconstructions) on all the subjects.

The *"per subject"* organization includes 23 folders, one for each volunteer, named according to the volunteers reference numbers (randomly assigned.
The *"per data"* organization consists of 5 folders, one for each class of data collected (high-frame rate videos with synchronous ECG, cf. Section 5; eye images, cf. Section 6; 3D Face scans, cf. Section 7; multispectral/hyperspectral imaging, cf. Section 8; images and videos for emotional testing, cf. Section 9.

Each "per subject" folder includes 5 sub-folders, one for each class of data acquired on that particular subject. In turn, "per class" folders include 23 subfolders, each one collecting that type of data for each of the volunteers.

The project partners had access to the data through HTTPS authenticated access. The data distribution required a high-speed connection, therefore we connected the server to a 1 GB backbone network which fitted with the expected client network.

10 CONCLUSIONS

In this document we have described the data collected during the SEMEOTICONS acquisition Campaign held in Pisa in May 2014. Data were collected as a result of a joint effort of partners involved in semeiotic modelling, methodological development and technological validation.

Following the campaign, a database including a huge set of data from 23 subjects was made available to the consortium. For each subject, it provides a complete medical and psychological characterisation focused on cardio-metabolic risk together with multimedia data acquired e from the face of volunteers. Data are currently being used to develop methods for extracting computational descriptor of facial signs possibly related to cardio-metabolic risk. In particular, the dataset usage has been reported in D3.1.1, D3.2.1, D3.3.1, and D6.1. In addition, data have used for preliminary testing the hypothesis underling the semeiotic model (See D1.1.2) and for the first technical validation report (D8.6.1).

In order to enrich the available dataset and enhance its statistical power, it has been agreed among partners to hold a second acquisition campaign. It has been planned to occur in Pisa in May 2015. The new data will be described in an addendum of this document, whose release is planned by the end of September 2015.

REFERENCES

[1]   http://www.medlab.eu/english/modules/ekgmodules/index.html
[2]   https://en.ids-imaging.com/store/produkte/kameras/ui-5240cp.html
[3]   http://www.itamar-medical.com/EndoPAT™/PAT_Technology
[4]   http://www.perimed-instruments.com/products
[5]   Strömberg T, Karlsson H, Fredriksson I, Nyström F, Larsson M. "Microcirculation assessment using an individualized model for diffuse reflectance spectroscopy and conventional laser Doppler flowmetry" Journal of Biomedical Optics 19,057002, 2014.
[6]   Fredriksson I, Burdakov O, Larsson M, Strömberg T. "Inverse Monte Carlo in a multilayered tissue model - merging diffuse reflectance spectroscopy and laser Doppler flowmetry" Journal of Biomedical Optics 18, 127004, 2013.
[7]   http://www.diagnoptics.com/age-reader/technology/

SEMEOTICONS reports

D1.1.2   Final Semeiotic model of cardio-metabolic risk.
D1.2   Physical models for multispectral imaging of the cardio-metabolic signs of the face.
D1.4   Validation protocol.
D2.1.1   Initial specification of system requirement and functionalities.
D3.1.1   MSI Hardware design and algorithms for monitoring endothelium function based on HSI.
D3.2.1   MSI Hardware design and algorithms for monitoring cholesterol level based on HSI.
D3.3.1   MSI Hardware design and algorithms for monitoring AGE accumulation based on HSI.
D4.1.1   3D geometric reconstruction subsystem.
D6.1   User's profiling tools.
D8.6.1   First report on technical validation.

APPENDIX I – SUBJECT INFORMATION FORM

Dear Sir/Madam,

Cardiovascular diseases (CVD) are the leading cause of morbidity and mortality worldwide. At present, the strategy of prevention, which attempts to modify some factors related to the origin of the disease, seems to be the only way to limit the epidemic growth of CVD.

Cardiovascular risk factors include obesity, physical inactivity, smoke, alcohol abuse, and abnormal lipid metabolism, hyper-glycaemia and arterial hypertension. Educational programs and lifestyle interventions represent effective tools to improve the individual risk profile and reduce the burden of CVD. Scientific evidence has shown a rapid decline in the risk for CVD mortality after individual or population-wide changes in diet and/or smoking and in general following an healthy life-style. In addition prevention has been shown to impact favourably on human wellbeing and health care costs. Unfortunately, less than one half of subjects at risk comply with the recommended lifestyles and achieves optimal values for blood pressure, lipids, and blood glucose in European surveys.

A rational prevention strategy alternative to intensive individual coaching by health professionals is the development of systems for self-learning and self-monitoring, designed to help people to change and maintain a healthy lifestyle. Human face has always been considered a mirror of emotions and mood and a reflection of the healthy or unhealthy status of individuals. SEMEOTICONS project will exploit human face as an indicator of individual's health status and derive measurements and descriptors automatically evaluated by a computerized application to be applied for CVD prevention.

Study description

The Italian National Research Council is the sponsor of this study that will be performed at 3 centres in Europe to develop an interactive "smart mirror", easily deployable in normal-life settings, which will seamlessly integrate contactless sensors and a touch-screen interface for user's interactions and output visualization.

The resulting device will be a Wize Mirror, which will collect data from face images and breath and derive a virtual individual's model used to compute and trace the daily evolution of an individual's wellness index.

The study will collect these novel measurements and compare them with standard tests commonly used in clinical practice, to check that measurements are accurate and reliable to identify changes in a subject's wellness.

The study will enrol approximately 80 subjects in Europe, 60 in Italy and 25 in France. The first 20 subjects will contribute to the development of the "Wize Mirror" and will be examined only once. In the remaining 65 subjects the effectiveness of the Wize Mirror in detecting changes in the wellness state associated with the adoption of a healthy lifestyle will be evaluated by repeating assessment after 3 months.

The study will last overall 3 years, the maximal duration of study period for each subjects is 3 months.

What you will be asked to do

Participants will undergo at recruitment medical visits and tests, which will require approximately half a day. You will be fasting since at least 8 hours. Assessments include standard clinical evaluations: medical history, physical examination, blood sampling, self administered questionnaires on food intake and physical activity, psychological profile, an electrocardiogram, measurement of body composition using pletismography, a vascular reactivity study and novel measurements to develop the Wize Mirror.

The pletismographic measurement of body composition is based on air volume displacement caused by subject entry into a closed chamber approximately 80 x 130 cm. You will have to stay within the chamber for 1 minute breathing normally into a tube.

The vascular reactivity study is performed in a room at controlled temperature (21-25°C) in the supine position by applying two finger-mounted probes, similar to thimbles to the second finger of each arm. A blood pressure cuff is placed on one upper arm (study arm), while the contra-lateral arm serves as control (control arm). After a 10-min rest period, blood pressure is measured on control arm and then recording

starts. The blood pressure cuff on the study arm is inflated to 60 mm Hg above the maximal blood pressure value for 5 min to slow down blood flow to the hand and then deflated. This trick stimulates the internal coating of blood vessels to release nitric oxide, that vasodilates the vessels. This test allows assessing the functionality of the arteries.

Photos and videos of your face will be taken and changes in skin temperature and blood flow to the face will be measured after exposure to the natural light or a thermal lamp to warm up face temperature up to a maximum of 42° C.

No genetic data will be collected. No experimental intervention is foreseen. Medical specialists will counsel you on advised lifestyle changes according to routine good clinical practice.

You will come back in the centre every 15 days to realise the Wize Mirror measurements (face image acquisition, thermal stimulation, and gas sensor acquisition) and to complete the lifestyle habits diary. You will be fasting since at least 8 hours minimum. The visit will take approximately 2 hours. At the end of study (three months), you will come back in the centre to realise the same measurements that during the first visit. The volunteers participating in this study cannot simultaneously participate in another biomedical research.

Risks and benefits

Sometimes blood sampling may cause slight pain and bleeding or numbness at the site of needle-stick insertion. Cuff inflation for 5 minutes may cause numbness or slight pain in the arm.

The 42° degree maximal temperature used to warm the face is similar to the one of a Turkish bath or to sun exposure in the summer in our country.

Direct benefits for you consist in a deeper understanding of your cardiovascular risk profile and a tailored prescription of lifestyle changes by medical specialists. If pathology is discovered, you and your personal medicine doctor will be informed.

Your participation will allow physicians to better understand the potential value of this novel tool, the Wize Mirror, for a simple and effective prevention strategy.

We will transmit you the overall results of study.

Face images will be used only to develop research activities. Non-medical researchers will not have access to demographic data. An additional informed consent will be required for permission to reproduce the images in scientific papers.

Consent to be enrolled in the clinical study

Your participation to this study is voluntary. You do not have to participate if you do not wish so. You have the right to change your mind and leave the study at any time you wish. No explanation is needed for quitting; you will simply have to tell a member of the research team. Your decision will not have any consequence. If you wish, your primary care physician may be informed of your participation to the study. The study protocol and the contents of this information sheet have been submitted to and approved by the Institutional Ethics Committee of Istituto di Fisiologia Clinica del CNR Pisa (Comitato etico - Azienda Ospedaliero-Universitaria Pisana) and Milano (Azienda Ospedaliera Niguarda), and the legal authorities in FRANCE (Comité de Protection des Personnes Sud-Est IV and Agence Nationale de Sécurité du Médicament) to Centre de Recherche Nutrition Humaine. The doctor responsible for the study will ask you to sign and date the Informed Consent Form attached to this document before submitting you to any procedure or test to ensure that you have been clearly and thoroughly informed and consent freely to participation.

The original document signed by you will be kept in the archives of Istituto di Fisiologia Clinica del CNR (Pisa and Milano)/Centre de Recherche Nutrition Humaine (Lyon). You will be given a copy of the document for your files.

Your personal data will be treated according to National Privacy Regulations.

If you have any question concerning your rights as research subjects you may ask the doctor responsible for the study:

Dr. _____   phone number _____

APPENDIX II – DELIVERABLE QUALITY PLAN

Deliverable identification:
- Deliverable name: Description of SEMEOTICONS reference dataset
- Contract date of delivery: Month 18
- Responsible: CNR
- Work package: 1
- Task: 1.3
- Type: Report
- Status: PU

1. Description according to the DoW

This report will describe the generation of the reference data set that will be used from the activities of WPs 3, 4, 5 and 6. The reference data will also provide a first test-bed for semeiotic modeling in WP1 and will be used in technological validation in Task 8.6.1.

2. Planning of the activities

This report describes the reference dataset as generated following the acquisition campaign held in Pisa in May 2014 and the following activities needed to organize the database, setting up the needing storage and sharing facilities. The acquisition campaign was planned starting in January 2014 and was organized by all partners involved in WP1, 3, 4, 5 and 6. First of all, we outlined the experimental protocol aimed at defining the clinical and psychological characterization of subjects with a special focus on cardio-metabolic risk (see D1.4). Afterwards, we defined and carefully planned the acquisition of multimedia data (including images, video, signals and 3D scans of the face) of volunteers. The overall planning of the acquisition campaign is the result of an intense cooperative action of the involved partners, which started at the end of January 2014. It is worth noting that, other activities in the Project, in particular those in Task 1.2 *Data modeling and investigation,* and Task 2.1 *Specification of system requirements and functionalities* are strictly related to the work undertaken for the acquisition of the reference dataset.   This is of paramount importance so a so to optimize the experimental setup and the overall acquisition protocol, ensuring a successful completion of the campaign and satisfying the methodological and technological requirements of the project.

An additional activity to be mentioned here is related to the logistics of the acquisition campaign, which, among other things, required the arrangement of special facilities (including dedicated rooms) so as to allow the involved researchers a correct and safe execution of the experimental labour. In addition, following the acquisition campaign, data post-processing (including anonymization and consistency check) was planned along with the design and implementation of data storage and sharing services.

3. Risk Analysis and contingency plans

Risks related with the activity are mainly those of the experimental work during the acquisition campaign. The major risk that can be foreseen and the related planning are the summarized in the following table.

Risk	Contingency plan
Some volunteers can withdraw from the study reducing the statistical significance of the dataset.	We foresee that this is a very low risk because enrolled volunteers will be selected among highly motivated people. In any case, to counteract the possibility of subject dropout, we plan to enroll a number of subjects slightly larger than planned one.
Some of the experimental devices used	The experimental setup is largely redundant, for most measurements

for acquisition can fail and some data can be lacking	acquisition with two different devices is carried out. In any case, integration of measurements with additional data is planned and a second acquisition campaign will be organized in 2015.

4. Quality conformity

This report conforms to the Project's Quality Assurance Process (QAP) as specified in the report D11.3 "Report on Quality Assurance Process".

APPENDIX III - QUALITY REPORT

(Comments are highlighted in italics)

Version	Date	From	To	Actions
1.0	6/7/2015	Giuseppe Coppini	Stelios Louloudakis	Version 1.0 of D1.3 sent for quality review
1.0	8/7/2015	Stelios Louloudakis Antonis Miliarakis	Giuseppe Coppini	Minor comments
1.1	8/7/2015	Giuseppe Coppini	Stelios Louloudakis Antonis Miliarakis	Document is amended according to IQR comments
1.2	8/7/2015	Stelios Louloudakis Antonis Miliarakis	Giuseppe Coppini	Quality clearance released

1) QUALITY STANDARDS
1.a) General quality standards (referring to the way of presentation)
The used language and style is accurate enough and easy to understand. Used terms are appropriate
The document's layout is compliant with the standardized project template.

1.b) Comment on general quality:
-The structure is clear, logic, and in accordance with good practice for writing reports for the scientific community. The content is in accordance with scientific standards on how to describe, discuss and conclude with regard to the issues addressed. Tables and figures are used in a clear, understandable and relevant way. The abbreviations table is complete and clear.

1.c) Specific quality standards:
1.c.1) Objectives achieved.
The Work-Package 1 *has fully achieved its objectives and technical goals for the period*
1.c.2) Fulfilment of deliverable's objectives for the period
The deliverable's objectives have been fulfilled according to the current timing and the available information up to the time of writing. An extensive analysis of all of the available information and proposed methods (as described in related deliverables) has been performed, while there is a detailed description of the data acquisition and volunteers recruiting method followed by the consortium.

2) DESCRIPTION OF THE QUALITY PROCESS (INCLUDING INTERNAL REVIEW)

Date	Action
3/7/2015	Stelios Louloudakis and Antonis Miliarakis (FORTHNET) appointed as IQR
6/7/2015	Document sent to for quality review
8/7/2015	Comments (minor issues) sent to task leader
8/7/2015	IQR approves the amended document and releases quality clearance
9/7/2015	Version 1.1 of D1.3 is sent to PC
10/7/2015	PC approves the document and delivers it to PO

2.a) Description of the Quality Process

On July 3rd 2015 FORTHNET as partner not involved in the activity reported in D1.3 was asked to act as quality reviewer for this deliverable. The same day, Dr Stelios Louloudakis and Antonis Miliarakis of FORTHNET accepted to review D1.3.

On July 6th 2015, version 1.0 of the document was sent to IQR for evaluation. On July 8th quality reviewers forwarded their comments to the task leader suggesting minor changes. An accordingly revised document was produced (V1.1 of D1.3) and resubmitted to IQR that approved it for quality. On July 9th Version 1.1 of D1.3 was sent to Project Coordinator for final approval.