This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 611516

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

D1.3 - Description of SEMEOTICONS reference dataset - Addendum

Delivery Date	31st October 2015
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	15/09/2015	CNR	ToC circulation
0.5	25/09/2015	All contributors	Collection of partial contributes
1.0	29/09/2015	CNR	Document integrated
1.0	14/10/2015	CNR	Document sent to IQR
1.1	20/10/2015	CNR	Minor changes according to quality review
1.1	21/10/2015	CNR	Quality approved by IQR
1.1	22/10/2015	CNR	Document sent to Quality Assurance Manager (Project Coordinator)
2	25/10/2015	QAM	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, Maria Antonietta Pascali
- FORTH: Franco Chiarugi,
- LIU: Tomas Strömberg, Marcus Larsson
- UCLAN: Bogdan Matuszewsky
- NTNU: Lise Randeberg

TABLE OF CONTENT

ABBREVIATIONS AND ACRONYMS

2014-ACQC	2014 Acquisition Campaign (First Acquisition Campaign)
2015-ACQC	2015 Acquisition Campaign (Second Acquisition Campaign)
AGE	Advanced Glycosylated End products
AST	Aspartate AminoTransferase
BMI	Body Max Index
DoW	Description of Work
DQP	Deliverable Quality Plan
ECG	ElectroCardioGram
FINDRISC	Finnish Type 2 Diabetes Risk Score
FLI	Fatty Liver Index
GGT	Gamma-Glutamyl Transferase
HbA1c	Glycosylated Hemoglobin, Type A1C
HDL	High Density Lipoprotein
HOMA	Homeostasis Model Assessment index
HR	Heart Rate
HSCORE	Heart SCORE
HSI	Hyper-Spectral Imaging
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
LOT-R	Life Orientation Test-Revised
MSI	Multi-Spectral Imaging
NUMi	Numeracy Understanding in Medicine Instrument
PSS	Perceived Stress Scale
SF-12	Health Survey
SF12-MCS	Mental Health
SF12-PCS	Physical Health
UV	Ultraviolet
WM	Wize Mirror
WP	Work Package

TABLE OF FIGURES

TABLE LEGENDS

EXECUTIVE SUMMARY

This document integrates the content of deliverable *D1.3 - Description of SEMEOTICONS reference dataset* with description of the data acquired in May 2015 during the second SEMEOTICONS acquisition campaign (2015-ACQC). This campaign was planned at the end of Task 1.3 and was aiming at:

- extending the size of the population of the reference data-set;
- using an experimental setup very close to the multi-sensing system included in the Wize Mirror prototype.

Data usage was described in D1.3 relating to the 2014 acquisition campaign (2014-ACQC) and is focused at:

- developing and testing computational methods by: a) sample input data (video/images) as planned for the Wize Mirror implementation (D2.1.1, D2.1.2), and b) reference ground truth made available by independent measurements (WP3, 4 and 5),
- verifying the hypothesis formulated in the final version of the semeiotic model of cardio-metabolic risk (work done in WP1) so as to optimize the overall development of the Virtual Individual Model in WP6,
- driving the semantic integration of self-monitoring data (Task 6.2).

The campaign was held in Pisa in the last week of May 2015, and was participated by the same partners of 2014-ACQC. The experimental protocol is the same used in 2014-ACQC, with a few modifications motivated by the changes in the acquisition setup and the feedbacks from 2014-ACQC data analysis. All the work was defined and thoroughly planned in Task 1.3.

PURPOSE AND SCOPE OF THIS DOCUMENT

This document is an addendum to deliverable D1.3 and describes the data of SEMEOTICONS reference dataset collected during the second acquisition campaign held in Pisa in May 2015 (2015-ACQC) 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 by properly integrating the content of D1.3. Relation between different data and project activities is also outlined. The structure of this document reflects the structure of D1.3. Similarly, for each of the 26 volunteers involved in the second campaign we obtained 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 applies to all partners involved in modelling (i.e. partners involved 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 we summarize the 2015 acquisition campaign.
- In Section 2 we describe the general characteristic of the dataset population.
- In Section 3 the clinical and 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 describes the acquisition of still pictures of eye.
- Section 5 reports the acquisition of 3D scans of the face.
- In Section 6 describes the acquisition setup and data collected by Hyperspectral/Multispectral imaging for endothelial function assessment, skin cholesterol and skin AGE evaluation.
- In Section 7 we describe the acquisition of images and movies during emotional stimulation.
- In Section 8 the Wize Sniffer acquisitions are summarized.
- General conclusions are drawn in Section 9.

Data storage and sharing facilities have been already reported in Section 9 of D1.3.

1 INTRODUCTION

According to the Project work-plan the SEMEOTICONS reference dataset is a core tool for almost all RTD activities including modelling, methodological and technological development. In fact, it provides medical data useful to verify the semeiotic model of cardio-metabolic risk. In addition, multimodal data are crucial to develop methods and algorithms in WP3, 4, and 5. The reference dataset is also expected to drive the design and implementation of the Virtual Individual Model (VIM) and the computation of the Wellness Index in WP6. Finally, the acquired data impact directly on technological work for *Wize Mirror* manufacturing. In fact, the reference dataset offers a common base for technological validation tasks. Extended details on all these aspects can be found both in the Description of Work and in previous reports.

The first version of the reference dataset is described in D1.3 and is a major product of the first acquisition campaign held in Pisa in May 2014 (2014-ACQC). After annualizing data so obtained according to both the medical and the technological viewpoint, and taking the progress of the project into account, it was agreed to extend the reference dataset in order to:

- Improve the statistical power of the sample with respect to target population. Modelling activities are based on the exploitation of medical knowledge applied to the available data. Therefore, a more numerous and representative dataset is expected to rise the statistical significance of VIM parameter estimation. In addition, the modeling of the temporal behavior of the VIM would be improved by using, for each subject, more observations taken at different times.
- Acquire additional data to cope with the progress of research activity in WP3-5. For example, the use of a fan heater had demonstrated to produce a vasodilatation of skin microcirculation much more effective than the previously used infrared lamps.
- Availability of the first *Wize Mirror* prototype released at Month 18 (MS3). The prototype is equipped with a suite of sensors tuned according to requirements in D2.1.1 and D2.1.2. while a large part of the data acquired in 2014-ACQC were obtained with a laboratory setup.

All these facts lead the consortium to implement the second acquisition campaign to be held in Pisa in the last week of May 2015 (2015-ACQC). It is worth mentioning that the potential need of a second acquisition campaign has emerged and discussed among partners since the early phases of the project and it was confirmed at Heraklion plenary meeting (November 2014).

Following discussion among research partners, it was agreed to extend the reference dataset sample by including in the study:

a) the subjects enrolled in 2014-ACQC;
b) a new set of volunteers.

In this way, after the 2015-ACQC, the reference dataset comprises: a) a group of persons (G1) with two observations, t0 (baseline) and t1, which would provide follow-up data useful in modelling activities, b) an additional group (G2) of volunteers with a single set of measurements taken at t0 (baseline).

The experimental protocol (see D1.4) adopted in 2015-ACQC is the same used in the first campaign with minor changes and optimizations that are mainly related to the usage of an updated acquisition setup as clarified in the remaining sections of this report.

On average, each volunteer had to be available for data acquisition for 1 - 2 hours for 2 times. Face images/videos acquisition took about 90 min. Similarly to 2014-ACQC 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), and the multimedia data obtained from each subject (Sections 4-8). The facilities used for storing and sharing the data in the consortium were described in Section 9 of D1.3.

2 DATASET POPULATION

At the end of March 2015, the medical researchers of SEMOTICONS contacted the 23 volunteers of 2014-ACQC (group G1(t0)) to assess their willingness to take part in the second campaign. Fourteen of them agreed to participate and formed the group G1(t1). An additional set of 12 subjects, forming the group G2(t0), was enrolled in the study according to the inclusion/exclusion criteria already defined in D1.4. A total of 26 volunteers entered the 2015-ACQC.

As detailed in next sections, the volunteers underwent the experimental protocol that is fully described in *D1.4 Validation protocol* and is summarized in Section 2.1 of D1.3 related to 2014-ACQC.

1.1. SUMMARY OF EXPERIMENTAL PROTOCOL

Each subject was characterized by clinicians and psychologists according to the following items:

- Clinical Assessment (Section 3.1);
- Psychological and nutritional tests (Section 3.3)

In addition, in the joint phase o the campaign (May 22-29) the following multimedia data were obtained from each subject:

- Eye images (Section 4)
- 3D Face scans (Section 5)
- MSI/HIS images (Section 6)
- Images and movies during emotional testing (Section 7).
- Wize Sniffer testing (Section 8).

Acquisition of multimedia data required updating the experimental setup by the partners involved in 2015-ACQC.

1.2. POPULATION DEMOGRAPHICS

The subjects in group G1(t1) included 11 males and 3 females. Age was between 32 and 62 years, mean age 48.1 years (SD 9.9 years). Female mean age was 57 years (SD 4.3 years), males mean age 45.7 was 45.7 (SD 9.7 years).
The subjects in group G2(t0) included 8 males and 4 females. Age was between 29 and 61 years, mean age 46.0 years (SD 9.42 years). Female mean age was 46.0 years (SD 13.1 years), males mean age was 46.1 years (SD 8.1 years).
The age distribution of the overall 20165-ACQC sample (G1(t1) and G2(t0))n is summarized in Figure 1

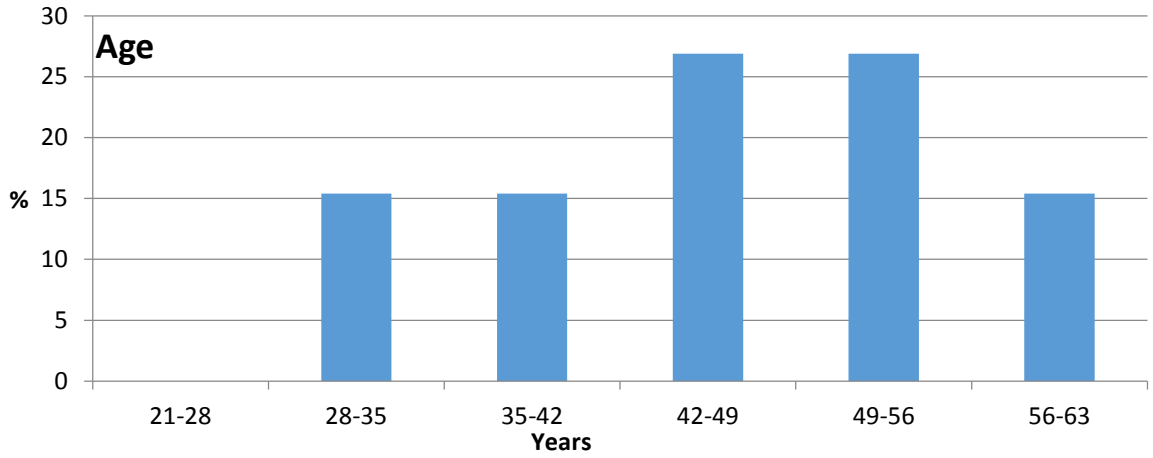

Figure 1. Histogram of age distribution in the second acquisition campaign sample.

In Figure 2 we summarize the major demographic features of the overall group of volunteers in 2015-ACQ. The 2015-ACQC G2 group includes three African ethnicity males (two Ethiopic and one Ivorian), all the other being Caucasian. The G1 group, as all the subject in 2014-ACQC, includes only Caucasians ethnicity persons.

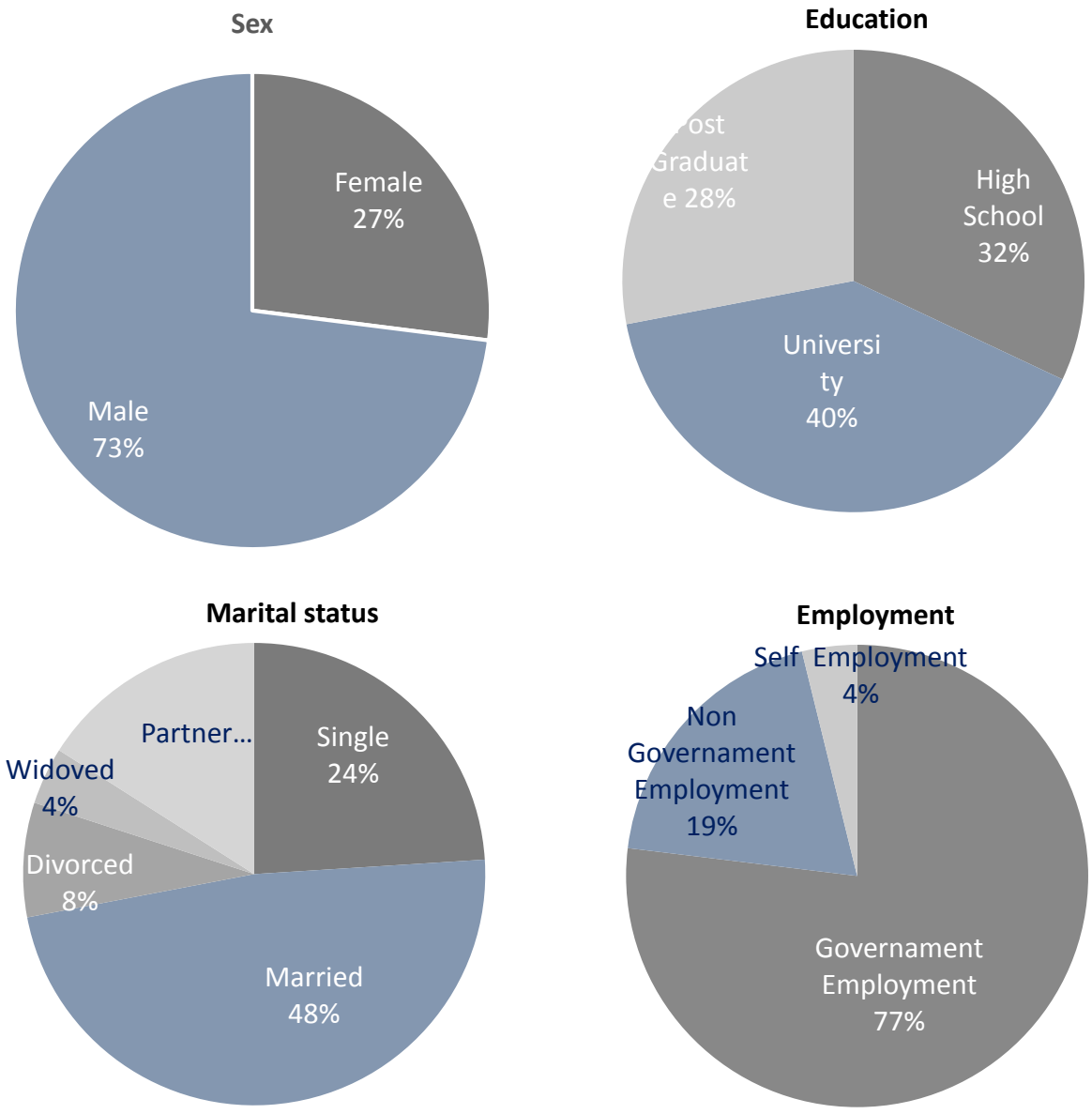

Figure 2. Main demographic features of 2015-ACQC sample.

3 CLINICAL AND PSYCHOLOGICAL CHARACTERIZATION OF VOLUNTEERS

3.1 CLINICAL ASSESSMENTS

Descriptive statistics (mean, median and standard deviation) of the measured parameters pertaining the clinical evaluation of the overall set of volunteers (G1(t0)+G2(t0)) are provided in Table 1.

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

Feature	Mean	Median	SD
Height cm	170.85	170.50	8.26
Weight Kg	75.82	73.10	22.88
BMI	25.70	25.75	6.36
Waist Circumf cm	90.63	89.50	19.73
HIP Circumf cm	99.96	99.00	15.37
Lean Mass % (imped balance)	77.00	77.10	5.93
Fat Mass % (Imped balance)	22.96	22.90	5.98
Lean Mass % BodPod	73.26	73.65	8.47
Fat Mass % BodPod	26.7	26.4	8.5
Thoracic Gas Volume l	3.64	3.48	.69
Energy Expenditure Kcal/day	1733.65	1679.50	330.38
Delta% of EE predicted	12.92	13.30	4.59
Respiratory Quotient	.81	.81	.08
VE l/min	28.10	27.10	5.86
Oxygen Saturation %	98.08	98.00	0.81
Heart Rate (bpm)	62.19	60.50	8.87
Systolic Pressure mmHg	127.19	130.00	8.98
Diastolic Pressure mmHg	76.77	80.00	6.59
RHI	1.59	1.62	.39
lnRHI	.49	.51	.22
AI %	7.60	3.00	18.65
AI@75 %	1.2	4.0	19.9
AGE_AF1	2.14	2.10	.46
HB gr/dl	14.99	15.10	1.17
RBC ul	5204615	5105000	527031
WBC ul	5975	6290	1076
Platelets ul	255731	248500	58939
Bilirubine mg/dl	.92	.86	.28
GOT mU/ml	22.9	23.0	5.1
GPT mU/ml	23.4	20.0	12.5
GGT U/L	28.0	25.5	13.2
Cholestorol mg/dl	240.7	234.5	43.6
HDL mg/dl	57.54	53.50	17.80
LDL mg/dl	160.85	153.00	36.99
Triglicerides mg/dl	109.96	89.00	68.23
Glucose mg/dl	90.81	89.00	9.14
HBa1c mmol/mol	37.23	37.00	3.36
INSULIN pmol/l	11.09	8.97	5.25
Creatinine mg/dl	0.94	0.95	0.14

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

Risk factor	Presence	Frequency	%
Hypertension	No	23	88.46
	Yes	3	11.54
Hypercholesterolemia	No	12	46.15
	Yes	14	53.85
Tobacco use	No	21	80.77
	Yes	5	19.23
BMI	Normal weight	12	46.15
	Overweight	11	42.31
	Obesity	3	11.54

In Table 2 we show a summary of the presence of hypertension, hypercholesterolemia, the tobacco usage (smokers) and the fraction of overweight and obese subjects, according to BMI value. Frequency and percentage are reported.

3.1.1 Risk score calculation

As in 2014-ACQC, for each subject the values of HSCORE, HOMA index, FLI and FINDIRISK were computed. In Table Table 3, we give central values (mean, median) and dispersion (SD) of these scores for the overall group of subjects at t0.

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

Score	Mean	Median	SD
HSCORE	1.04	0.83	1.13
HOMA Index	2.42	1.97	1.39
FLI	35.23	32.00	27.29
FINDRISC	6.58	6.00	3.53

The histograms plotted Figure 3 provide and overall view of the risk score in the studied population. As it can be expected, 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.

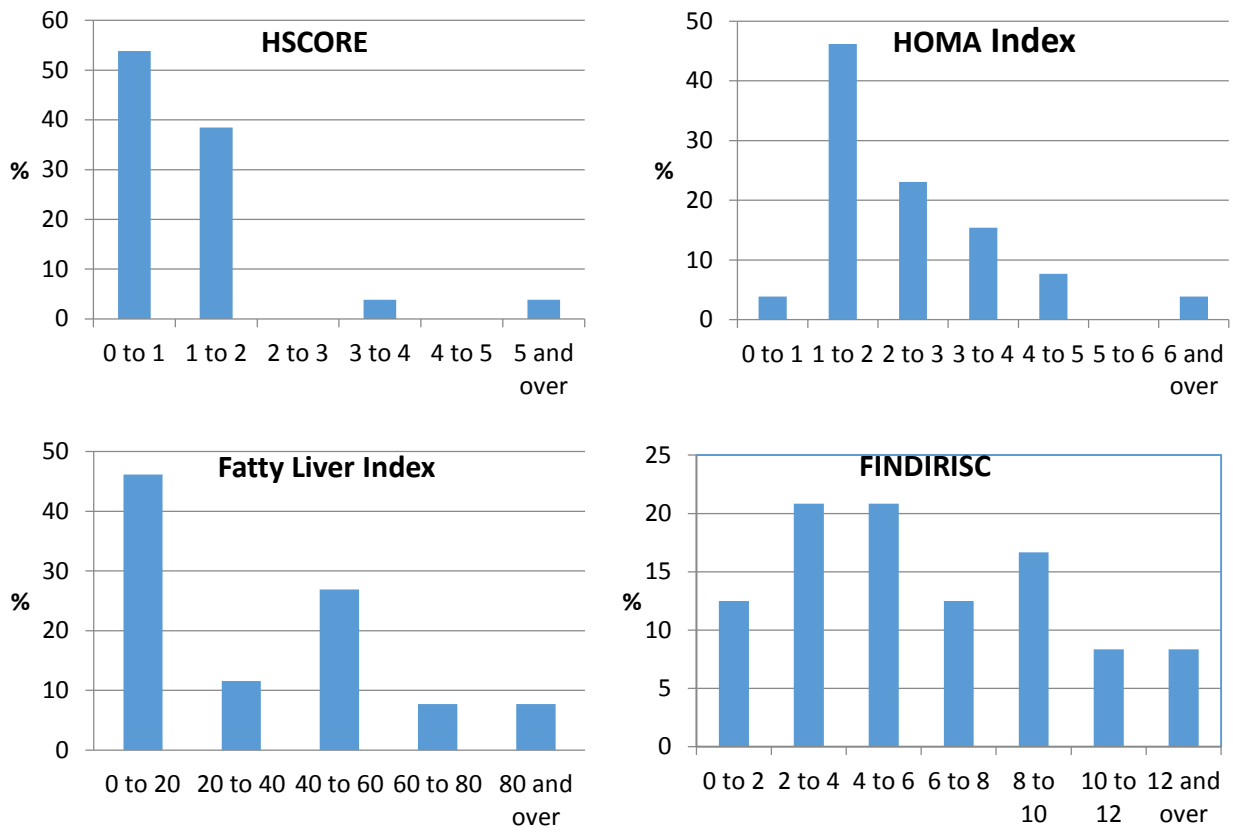

Figure 3. Distribution of risk scores in the population of the second acquisition campaign.

3.2 PSYCHOLOGICAL, EDUCATION AND NUTRITION QUESTIONNAIRES

As to psychological and behavioural characterization it was decided to dismiss the Numeracy Understanding in Medicine Instrument (NUMI). In fact, in the first campaign, NUMI was rated too complex, time-consuming, and just acceptable by most subjects. All the other questionnaires were retained in 2015-ACQC. In order to optimize acquisition timing and following discussion among research partners, it was agreed to administer these questionnaires only at t0 volunteers. Therefore, only G2 group filled these questionnaires.
In addition, the activities ongoing in WP6 (mostly T6.3) have suggested the adoption a new set of behavioural questionnaires for an almost direct implementation in the *Wize Mirror* (AUDIT-C relating to alcohol consumption, Fagestrom for smoke habits, and IPAQ for physical exercise) that are characterized by shortness and simplicity and are expected to produce valuable information for user guidance. Therefore, AUDIT-C, Fagestrom, and IPAQ were administered to all subjects (G1(t1)+G2(t0)) in 2015-ACQC.

Table 4. Behavioral and psychological features of SEMEOTICONS subjects at t0

Feature		Freq	%	Cum %
Tobacco use	Non-smoker	9	90.00	90.00
	Smoker but not every day	0	0.00	0.00
	Every day smoker	1	10.00	100
Alcohol use	Non-drinker	0	0.00	0.00
	Drinking in the last year	2	18.18	18.18
	Drinking in the last month	9	81.82	100
Nutrition	No risk	2	20.00	20.00

	Medium risk	3	30.00	50.00
	High risk	5	50.00	100
Physical activity	Low level	6	50.00	50.00
	Medium level	2	16.67	66.67
	High level	4	33.33	100
Anxiety	No anxiety	8	72.73	72.73
	Moderate anxiety	2	18.18	90.91
	Severe/moderate anxiety	1	9.09	100
Depression	No depression	10	83.33	83.33
	Mild depression	2	16.67	100
Stress	No stress	7	58.3	58.3
	Stress	5	41.7	100
Insomnia	No insomnia	12	100.00	100.00
	Sub threshold insomnia	0	0.00	100.00
	Moderate severity clinic insomnia	0	0.00	100
Motivation to change - physical activity	Pre-contemplation	1	8.33	8.33
	Contemplation	6	50.00	58.30
	Determination	3	25.00	83.30
	Action	1	8.33	91.60
	Maintenance	1	8.33	100.00
Motivation to change - nutrition	Pre-contemplation	2	16.67	16.67
	Contemplation	4	33.33	50.00
	Determination	2	16.67	66.70
	Action	1	8.33	75.00
	Maintenance	3	25.00	100.00

Table 5. Score about perceived wellness: SF12 scores for physical and mental health, Perceived Stress and LOT-R score (data for G2 group)

Score	Mean	Median	SD
Physical health (SF12 - PCS)	56.51	56.65	3.00
Mental health (SF12 - MCS)	45.31	46.15	8.37
Perceived Stress Scale	13.00	12.00	7.50
Optimism (LOT R)	21.42	23.00	5.38

Table 6. Results from the added pshychological and behavioural questionnaires for the entire sample of 2015- ACQC (G1(t1) and G2(t0) groups)

Feature		Freq	%	Cum %
Audit-C	Negative	21	80.80	80.80
	Positive	5	19.20	100
Fagerstrom (only for smokers)	Low dependence	4	80.00	80.00
	Low to moderate dependence	1	20.00	100
IPAQ	Low Physical Activity	7	26.90	26.90
	Moderate Physical Activity	19	73.10	100

4 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 risk scores (see D3.4.1). 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.

4.1 ACQUISITION SETUP

Following setup was used:

- Camera: CANON 5D mark II; Lens: CANON MACRO EF 100mm
- 4 LED light sources to provide diffused illumination
- Black screen around the subject to minimize reflections in the eye.

A professional photographer supported the study of the best set-up conditions. The acquisition system is shown in the panels of Figure 4.

4.2 DESCRIPTION OF DATA

For each of the **26** volunteers, several images of the same eye were captured. The images were saved in both raw and JPEG format (see Figure 4).

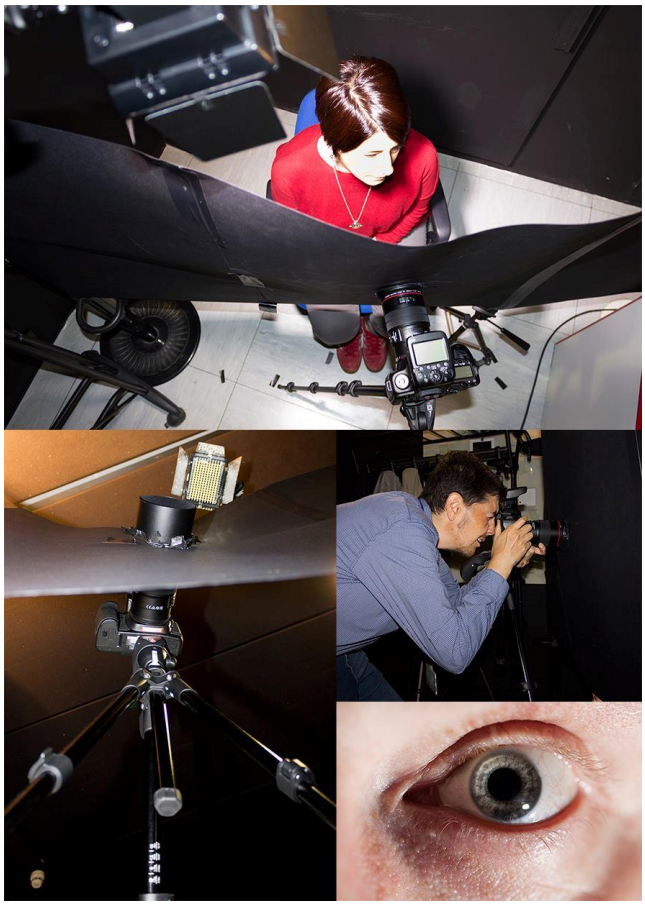

Figure 4. Eye Images acquisition setup with a sample image on the bottom right panel.

5 3D FACE SCANS

The acquisition of 3D faces relates to WP 4 "3D models construction and characterization". The 3D reconstructions are used for bio-morphometric data analysis (Task 4.3). The collected data also facilitates evaluation of the inexpensive 3D scanner developed specifically for the mirror (D4.1.1).

5.1 ACQUISITION SET-UP

Two different 3D scanners where used in the acquisition: (i) commercial Artec Eva, and (ii) inexpensive 3D scanner developed specifically for the *Wize Mirror* (WM). The 3D scans captured by the Artec scanner are treated as the reference data. The *Wize Mirror* scanner was used in two acquisition scenarios: (i) with head rotation and (ii) rotated full body. This was done to assess possible distortions caused by the articulation of the neck during the data acquisition on the WM.

5.2 DESCRIPTION OF DATA

The data from the Artec scanner is stored in Artec proprietary format as well as .ply file format. The data from the WZ scanner is stored in a raw binary format as well as .ply data file. On average the data from the Artec scanner contains approximately 200k polygons and 100k vertices, whereas the data from the WM scanner 40k polygons and 20k vertices

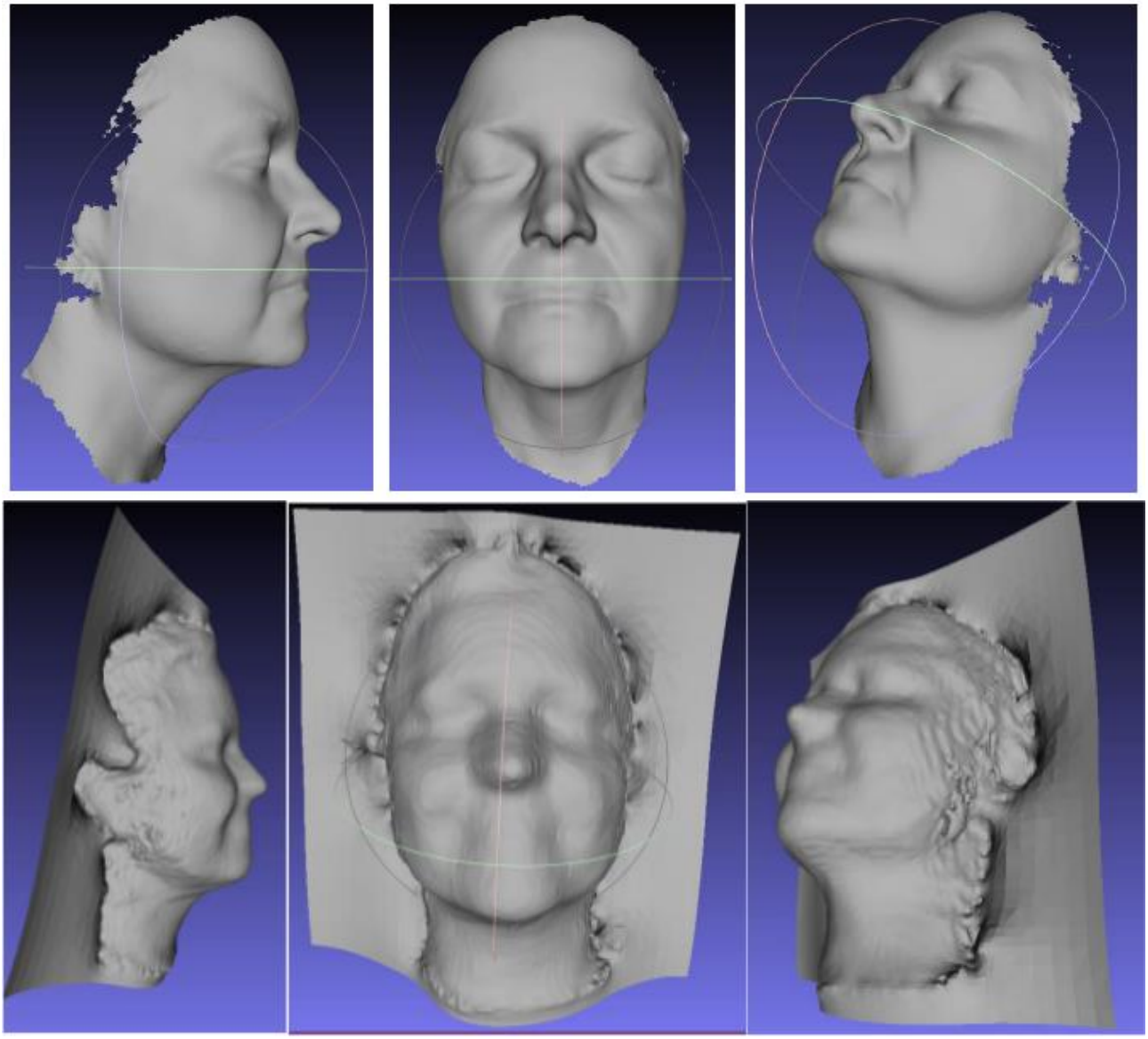

Figure 5. Samples reconstructions from Artec (top panels) and WM scanner (bottom panels)

6 MULTISPECTRAL / HYPERSPECTRAL IMAGING

6.1 MSI AND HSI SETUP FOR ENDOTHELIAL FUNCTION, AGE ACCUMULATION AND CHOLESTEROL LEVEL ASSESSEMENT

For each subject we acquired both hyperspectral and multispectral images. Two different Hyperspectral imagers were used. A first set of hyperspectral images was acquired by LIU team with an imager consisting of a camera equipped with a LCTF filter (see also D1.3), 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. These images are mainly used to test and tune methods for blood oxygenation and blood amount assessment.
Another set of hyperspectral images was acquired by NTNU group with the imager specifically developed in SEMEOTICONS (see D3.2.1 and D8.6.1) and based on a camera couple to an AOTF filter (HIS-440C with 18 mm lens, Gooch&Housego, UK) with a ring-LED white light source (see Figure 6, left panel). The spectral range was 446 – 702 nm with 4 nm step. The major usage of these images is the development and test of methods for skin cholesterol evaluation.

Multispectral images were obtained by the 5-camera MSI developed in SEMEOTICONS and included in the Wize Mirror prototype (See D3.3.1, D3.3.1 and D8.3), see Figure 6, middle panel. The related images are being used to test and tune methods and algorithms for the endothelial function, skin cholesterol, and skin AGE assessment by multispectral imaging.

Thermal heating for endothelial function assessment was based on a computer controlled fan heater (see Figure 6, right panel) equipped an IR thermometer (see D3.1.1). This heater replaces the IR lamps already tested in 2015-ACQC that revealed ineffective in inducing thermal vasodilatation.

Besides the references provided from clinical evaluation of subjects (see Section 3.1), LASCA perfusion images where acquired, additional evaluation of endothelial function was also derived by LDFRS.

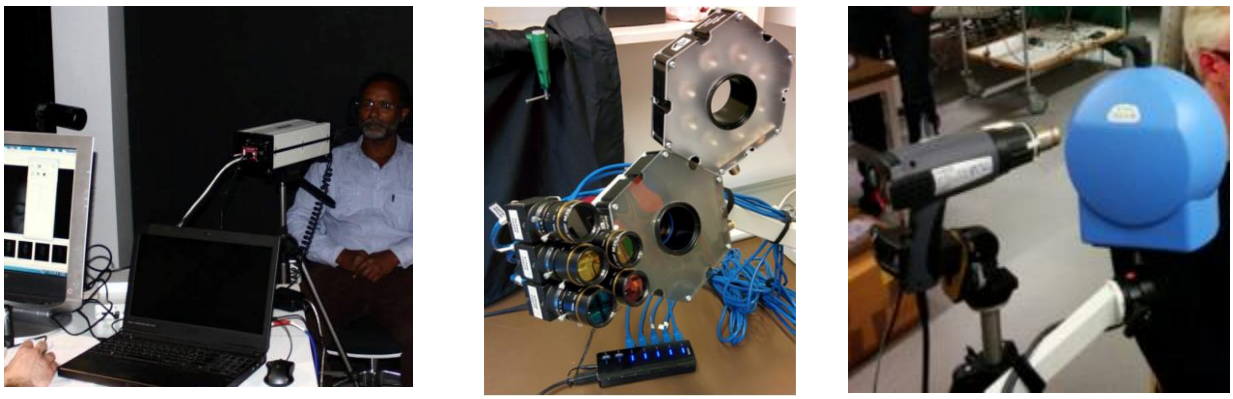

Figure 6. Left: NTNU hyperspectral imager; middle: 5-cameras MSI system; right: fan heater and LASCA imager.

6.2 SAMPLE IMAGES USING THE MSI HARDWARE FOR AGE ACCUMULATION AND ENDOTHELIAL FUNCTION ASSESSMENT

During the 2015-ACQC two cameras of the 5-camera MSI system (wavelength bands: 355nm and 475nm) was used for remotely quantifying the AGE product concentration in skin (Figure 7). One set of computer controllable UV LEDs were used to stimulate skin autofluorescence.

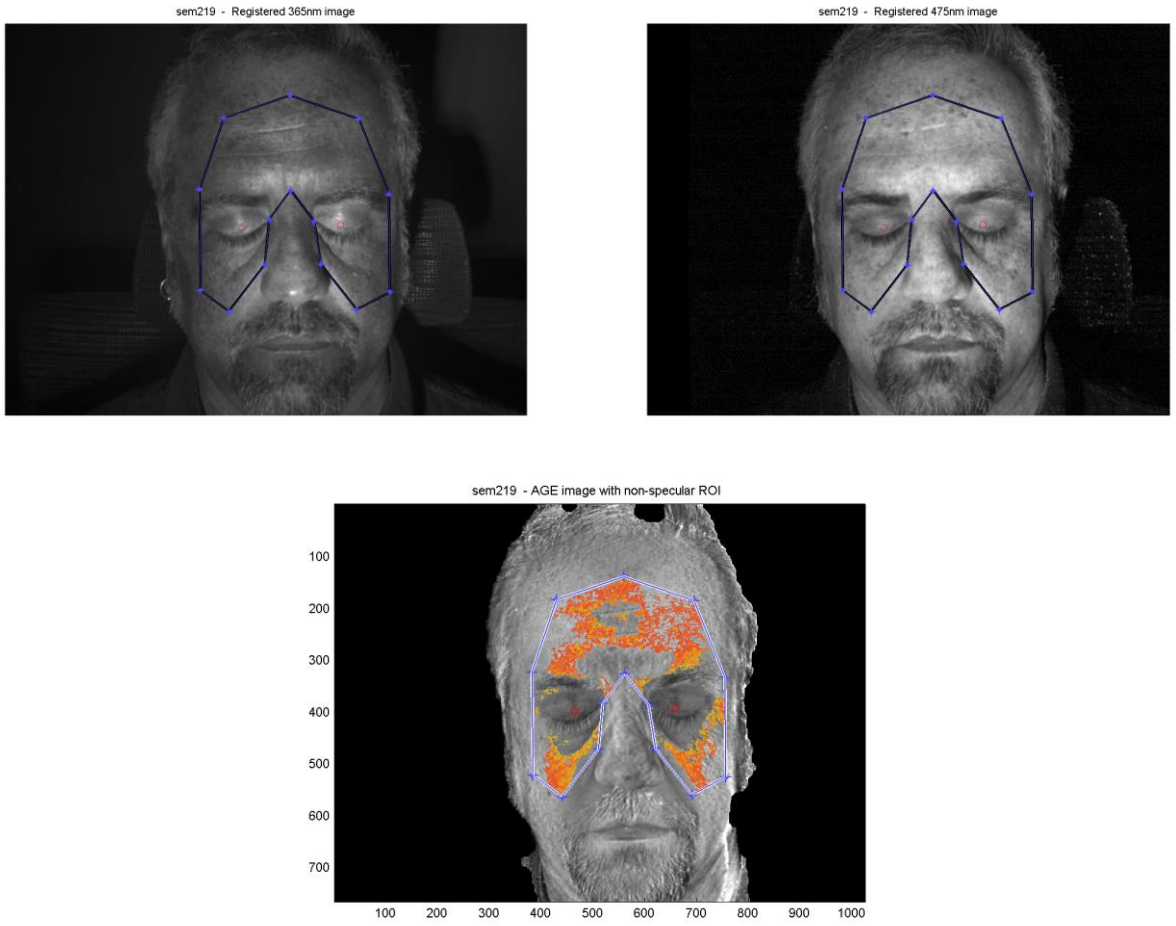

Figure 7. Upper left figure shows an example of a diffuse reflectance MSI image captured with a short pass optical filter (<400nm) under 365nm UV LED illumination. Upper right figure shows the UV induced auto fluorescence MSI image captured using an optical band pass filter at 475± 25nm. The lower figure shows the AGE image calculated as the ratio between the auto fluorescence image (upper right image; 475nm) and the diffuse reflectance image (upper left image; 365nm). The marked pixels are those within the mask that are without specular reflections and outside the eye region. The AGE concentration is calculated as the mean value from those pixels.

As to endothelial function assessment, a typical MSI recording, from the 2015 campaign is presented in Figure 8. The upper panels show a raw image from one of the four cameras, before facial skin heating (left) and after 9 min of facial skin heating using heated air from the computer controlled fan as provocation. Overlaid in the upper panels are the region of interest ROI in color, where the color represents calculated oxygen saturation S_{O2}. The lower panels show the calculated concentration of blood f_{RBC} pixel by pixel, but no raw image.

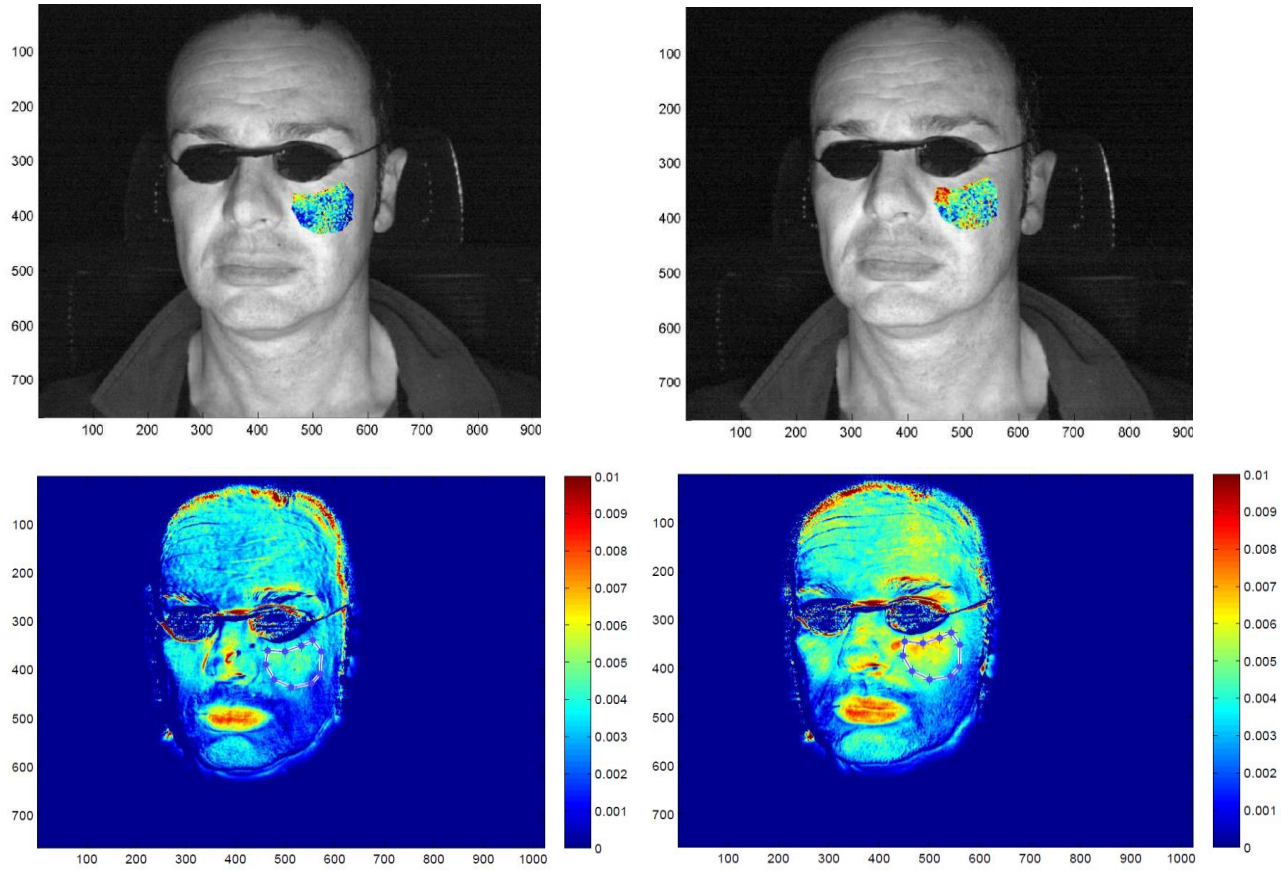

Figure 8. Typical MSI recording with ROI and calculated SO2 and fRBC at baseline (upper and lower left) and 9-10 min after local heating (upper and lower right). Upper panels are recordings by one camera with SO2 in ROI overlaid, lower panels are fRBC calculated for the whole image.

6.3 HIS FOR SKIN CHOLESTEROL ASSESSEMENT

Hyperspectral images were recorded before performing any provocations (UV irradiation, heating) thus resulting in baseline images. All images were converted to reflectance images using white standard calibration (at correct working distance) prior the analysis.

An image from a hyperspectral camera reflects the total amount of photons hitting the detector surface during the integration time. The data consist of a cube where the spatial coordinates are the two of the axes, and the spectral information is the third. In Figure 9, some of the wavelengths measured from a volunteer in 2015-ACQC are shown.

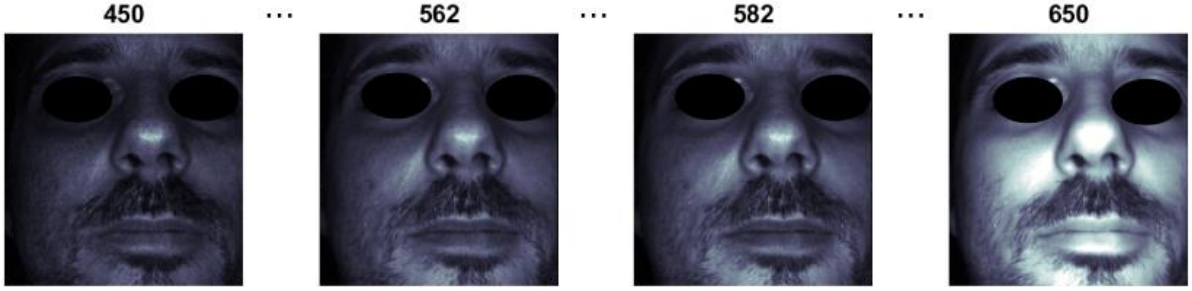

Figure 9. Some of the wavelength bands measured using the NTNU hyperspectral system from of volunteers in 2015-ACQC

During 2015-ACQC, additional spectral data, useful for algorithm development and tuning, were also acquired by using a spectrometer probe. Reflectance spectra were also measured using an integrating sphere (ISP-REF, Ocean Optics, FL, USA) and a spectrometer (SD1000, Ocean Optics, FL, USA) immediately after a HSI measurement for each volunteer. Three sites on volunteer's face were included: left cheek, right cheek and forehead. The resulting spectra included spectral range 534 – 1100 nm with 0.367 nm resolution.

7 IMAGES AND MOVIES DURING EMOTIONAL TESTING

7.1 ACQUISITION SETUP

The experimental procedure is aimed to record videos, conveying meaningful information to detect anxiety / stress / fatigue states from facial video recordings (see also D5.3.1). The participants were seated in front of a computer monitor while the camera was 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 a color video camera provided by FORTH.
The same camera and lens selected for the acquisition of videos, from which to compute the facial signs related to stress, anxiety and fatigue in the first prototype of the WM, were used in this experimental procedure. More specifically, the camera was the Point Grey Grasshopper 3 GS3-U3-41C6C-C USB with 1 x fixed focal length lens, FL = 16mm. This camera was not available at the time of the first acquisition campaign (Pisa I experiments) and it is worth noting that it allows higher space and time resolution in comparison with the camera used in Pisa I experiments.
The entire setup is shown in Figure 10.

Figure 10. The setup of the 2015-ACQC.

Other significant differences and improvements compared to the experiments performed in the first acquisition campaign were the use of the maximum space and time resolution of the new video camera compatibly with the maximum bandwidth of the overall acquisition and storage equipment (the bottleneck was the SSD writing channel, which was not fast enough in order to exploit the maximum allowable space and time resolution (2048 x 2048 at 90 fps)). Thus, the adopted space and time resolution (1600 x 1216 at 90 fps) in the Pisa II experiments was lower than the maximum allowable by the new video camera, but significantly higher than the one used in the first acquisition campaign. At the same time a polygraph was used for recording in parallel with the videos some relevant vital signs such as ECG and EMG, for the construction of a more reliable ground truth. Such a polygraph was not available in Pisa I experiments.

7.2 DESCRIPTION OF DATA

In this experiment, 26 volunteers took part in. From them, 24 subjects performed all stress/anxiety experiment tasks.
In Figure 11 a frame from a video recording during the Pisa II experiments is shown.

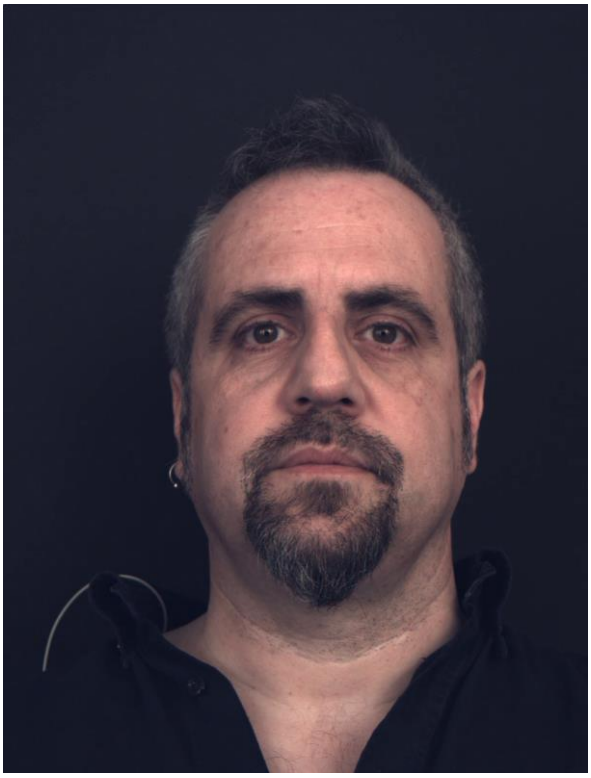

Figure 11. Sample of video acquired during 2015-ACQC.

The raw videos were collected and coded/compressed to Moving Picture Experts Group 4 (MPEG-4) (.mp4 file) in order to be convenient for further analysis. The coding was performed in 15000kbps so the size of each video is about 110MB/min.
The whole procedure aimed at inducing stress/anxiety states to participants and investigating these states. Therefore, volunteers were asked to describe themselves in a foreign language, to simulate a state of stress/anxiety/fatigue, to watch stressful audiovisual stimuli and to perform difficult cognitive tasks (both shown on the computer monitor in front of them). Stimuli included images having stressful content such as human pain, violence against women and men, armed children, drugs usage, etc. As regards videos, there was a relaxing video from an exotic island and two stressful videos of a drowning and a probable death caused by someone being buried alive. 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 self-assessment of the experienced feeling was stored as part of the dataset.
The overall procedure is depicted in Table 7.

Table 7. Experimental procedure of Pisa II Acquisition campaign.

1	Reference period (1 min)
2	Self-describing (1 min)
3	Mimicking neutral/rest state (1 min)
4	Mimicking anxiety state (1 min)
5	Mimicking stress state (1 min)

6	Mimicking fatigue state (1 min)
7	Sequential non calming images presentation (2 min)
8	Mental Task (Stroop test) (2 min)
9	Recording with blank monitor (1min)
10	Recording viewing calming videos (2 min)
11	Recording viewing adventure film (2 min)
12	Recording viewing adventure film (2 min)

8 WIZE SNIFFER

8.1 EXPERIMENTAL PROTOCOL

During SEMEOTICONS second acquisition campaign we considered a population of 26 subjects. The Wize Sniffer test was made following individuals' habits. That explains the reason why there was not a particular measuring protocol: we asked the subjects simply to follow their daily habits.
Regarding the sampling procedure, we respected the one implemented for the first technical validation (D8.6.1). As previously described, in practice three methods of sampling are generally used: "alveolar (end-tidal) sampling" (which corresponds to the plateau of the CO2 curve, that is the maximum value of exhaled CO2), "mixed expiratory air sampling" (which corresponds to a whole breath sample), "time-controlled sampling" (which corresponds to a part of exhaled air sampled after the start of expiration). Controlled alveolar sampling by means of expired CO2 concentrations is the method of choice if systemic volatile biomarkers are to be assessed, since only alveolar compounds are correlated to compounds in blood. The mixed expiratory air sampling, without controlled identification of the respiratory phases bears the risk of dilution with dead space air, but this method allows collecting also the compounds of exogenous origin. The last method shows large variations of compositions because of wide variations of individual dead space volumes and breathing maneuvers, so it is less used in clinical practice. For our purposes, the mixed expiratory air sampling method was chosen, since our interest was focused on endogenous biomarkers, as well as compounds of exogenous origin (such as ethanol, for example). In addition, since composition of single breaths may vary considerably among different individuals, because of different modes and depth of breathing, we asked the subjects to take a deep breath and to exhale into the Wize Sniffer's disposable mouthpiece emptying as much as possible their lungs. This precaution was taken in order to have breath samples that were as reproducible as possible.
The test took up to 2 minutes. For each subject details about smoking, alcohol intake, nutrition, and lifestyle are gathered from clinical, psychological and behavioral characterization (see Section 3).

8.2 EXPERIMENTAL SET-UP

Our experimental set-up consisted of the Wize Sniffer (whose description is reported in detail in the deliverables D3.5.1 and D3.5.2), disposable mouthpieces, and a personal computer. The WS data were collected making a request from the PC to the Telnet Server (port 23) implemented on Arduino Mega2560 board with Ethernet module. A TCP/IP Protocol was used. The software used to collect WS data was implemented in Matlab.

8.3 ACQUIRED DATA

The data collected during the SEMEOTICONS acquisition campaign consisted, for each sample, of the maximum voltage values output from the gas sensors (raw data).
They are reported in Table 8.

Table 8. Wize Sniffer raw data collected during 2015-ACQC. In the first column, the Subject ID; in the second column the exhaled volume (ml); from the third column to the tenth one, the gas sensors' maximum voltage output (Volt); in the last two columns the temperature and relative humidity values.

SubjID	V(ml)	TGS4161 (V)	TGS2620 (V)	TGS821 (V)	TGS2602 (V)	TGS2444 (V)	CTS_O2 (V)	CTS_CO2 (V)	TGS2442 (V)	TEMP (°C)	HUM (%)
215	1854	4.03	1	0.87	1.86	0.11	2.76	3.48	0.84	26.37	77.92
218	3170	4.02	1.23	1.33	1.45	0.11	2.45	3.88	0.84	26.64	78.31
211	2243	4.03	0.94	0.58	2.29	0.14	2.39	3.93	0.79	25.59	79.99
201	2101	4.02	0.64	0.31	1.6	0.13	2.39	3.93	0.82	25.8	83.36
207	1941	4.02	0.72	0.46	1.22	0.14	2.39	3.93	0.82	25.88	84.76
213	2326	4.03	1.39	1.58	1.93	0.12	2.35	3.95	0.81	26.38	84.6
208	1941	4.03	1.21	1.42	1.62	0.12	2.35	4	0.83	26.46	73.54
221	2400	4.05	1.48	1.73	1.86	0.12	2.35	4	0.84	26.73	82.21
220	2454	4.04	1.59	1.95	1.72	0.1	2.33	4.07	0.84	27.27	81.93
214	2157	4.04	1.01	0.78	2.02	0.1	2.33	4.07	0.8	26.38	70.45
206	1349	4.03	0.78	0.5	1.9	0.11	2.33	4.07	0.81	26.67	81.9
205	2116	4.04	0.81	0.54	1.67	0.1	2.33	4.07	0.81	26.43	73.54
223	2116	4.05	0.89	0.89	1.5	0.1	2.33	4.07	0.82	26.41	73.39
212	2608	4.04	0.72	0.32	2	0.09	2.33	4.07	0.82	26.8	80.43
216	2960	4.04	0.84	0.59	1.77	0.09	2.33	4.07	0.81	26.94	81.04
217	2289	4.04	1.31	1.43	1.94	0.09	2.33	4.07	0.82	27.22	84.14
203	2116	4.04	1.26	1.53	1.82	0.1	2.23	4.07	0.83	27.13	73.48
209	1165	4.03	0.72	0.42	1.69	0.05	2.31	4.04	0.79	27.35	54.79
202	3420	4.03	0.83	0.55	1.86	0.07	2.31	4.04	0.8	27.46	77.31
210	1880	4.03	1.2	1.26	1.6	0.07	2.87	4.04	0.8	27.71	75.46
225	1836	4.03	1.12	1.07	1.64	0.08	2.75	4.04	0.8	27.95	75.84
224	1172	4.03	0.71	0.58	1.65	0.07	2.75	4.04	0.82	27.56	40.84
204	2595	4.05	1.53	1.69	2.17	0.08	2.46	3.54	0.82	29.09	71.42
219	1610	4.04	1.21	1.39	1.65	0.14	2.46	3.71	0.77	26.69	68.32
226	3608	4.04	0.68	0.26	1.72	0.12	2.46	4.04	0.77	27.09	79.9
222	918.5	4.04	0.54	0.24	1.37	0.11	2.46	4.04	0.77	27.01	47.92

Data format: ASCII text
Size: 10-100Kbits for each sample

9 CONCLUSIONS

In this document we have described the data collected during the second acquisition campaign held in Pisa in May 2015. Data were produced as a result of a joint effort of partners involved in semeiotic modelling, methodological development and technological validation.

Merging the results of both 2014_ACQC and 2015_ACQC, SEMEOTICONS reference dataset now includes data and measurements (observations) from 35 individuals: 14 of them have two observations after one year from each other (t0 and t1 respectively), and 21 (9 from 2014-ACQC and 12 from 2015-ACQC), of them have a single baseline observation.

The acquisition protocol of previous campaign was kept, with optimization aimed to a) account for feedbacks from last year RTD activities, this includes, for example, the adoption of new behavioural questionnaires, and b) use a hardware and software setup very close the one of the Wize Mirror prototype.

For each observation of a subject, we are able to provide a complete medical and psychological characterisation focused on cardio-metabolic risk together with multimedia data acquired from volunteers. The dataset is shared among the partners and is used for ongoing RTD activities

REFERENCES

SEMEOTICONS reports:

D1.3 Description of SEMEOTICONS reference dataset
D1.4 Validation protocol.
D2.1.1 Initial specification of system requirement and functionalities.
D2.1.2 Revised 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.
D3.4.1 Eye image characterization
D3.5.1 Gas sensors for breath analysis
D3.5.2 Design and integration of the Wize Sniffer
D4.1.1 3D geometric reconstruction subsystem.
D5.3.1 Algorithms and methods for facial expression analysis and psycho-physical status evaluation
D8.3 Release of the first *Wize Mirror* Prototype
D8.6.1 First report on technical validation.

APPENDIX I – 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 2nd acquisition campaign held in Pisa in May 2015. The acquisition campaign was planned starting in January 2015 and was organized by all partners involved in WP1, 3, 4, 5 and 6. The experimental protocol is the same adopted in the 1st acquisition campaign (see D1.3) aimed at defining the clinical and psychological characterization of subjects with a special focus on cardio-metabolic risk (see D1.4). Minor changes to the protocol were introduced both to account for a) feed-back from year two project activities, and b) to utilize an experimental acquisition setup similar to the one embodied in the Wize Mirror Prototype (D2.1.2). The overall planning of the acquisition campaign is the result of an intense cooperative action of the involved partners, starting at the end of January 2015. This led 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. Data storage and sharing services already implemented for the 1st acquisition campaign are being use also for 2015 data.

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 for acquisition can fail and some data can be lacking	The experimental setup is largely redundant, for most measurements acquisition with two different devices is carried out.

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 II - QUALITY REPORT

Version	Date	From	To	Actions
1.0	14/10/2015	Giuseppe Coppini	Christine Assimakpoulo (IQR)	Version 1.0 of D1.3 sent for quality review to IQR
1.0	20/10/2015	Christine Assimakpoulo	Giuseppe Coppini	Minor comments
1.1	20/10/2015	Giuseppe Coppini	Christine Assimakpoulo	Document amended according to IQR
1.1	21/10/2015	Christine Assimakpoulo	Giuseppe Coppini	Quality clearance released

1) QUALITY STANDARDS

1.a) *General quality standards (referring to the way of presentation)*

- The language and the style are suitable for the deliverable. Overall the document is easy to follow.
- The terminology used is appropriate and easy to understand.
- The document's layout is compliant with the standardized project template

1.b) *Comment on general quality:*

The structure of the deliverable is clear, logic, and in accordance with its proposed content. The content is in accordance with scientific standards on how to describe, discuss and conclude with regard to the issues addressed.

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

Date	Action
12/09/2015	Christine Assimakpoulo (FORTHNET) appointed as IQR
14/10/2015	Document sent to IQR for quality review
20/10/2015	Comments (minor issues) sent to task leader
20/10/2015	Document amended following IQR comments
21/10/2015	IQR approves the amended document and releases quality clearance
22/10/2015	Version 1.1 of D1.3 is sent to PC
25/10/2015	PC approves the document and delivers it to PO

2.a) Description of the Quality Process

On September 9rd 2015, Christine Assimakpoulo by FORTHNET was appointed as internal quality reviewer for the addendum to D1.3.

On October 14th 2015, version 1.0 of the document was sent to IQR for evaluation. On October 20th reviewer forwarded her comments to the task leader signalling a few typos and minor changes. An accordingly revised document was produced (V1.1 of D1.3) and resubmitted to IQR that approved it for quality. On October 22nd Version 1.1 of D1.3 was sent to Project Coordinator for final approval.