International Journal on

Advances in Life Sciences

2018 vol. 10 nr. 1&2

The *International Journal on Advances in Life Sciences* is published by IARIA.
ISSN: 1942-2660
journals site: http://www.iariajournals.org
contact: petre@iaria.org

International Journal on Advances in Life Sciences
Volume 10, Number 1 & 2, 2018

Eiko Yoneki, University of Cambridge, UK

International Journal on Advances in Life Sciences
Volume 10, Numbers 1 & 2, 2018

CONTENTS

A Holistic Approach for Forecasting Medical Equipment Risks Using Monte Carlo Simulation

Sahar Ismail, Hassan Nehme, and Mohamad Hajj-Hassan
Dept. of Biomedical Engineering
Lebanese International University
Beirut, Lebanon
Email: sahar.ismail93@gmail.com
Email: hassan.nehme@live.fr
Email: mohamad.hajjhassan@liu.edu.lb

Bassam Hussein
Department of Industrial Engineering
Lebanese International University
Beirut, Lebanon
Email: bassam.hussein@liu.edu.lb

Hassan M. Khachfe
Center for Quality Assurance and Scientific Research
Lebanese International University
Beirut, Lebanon
Email: hassan.khachfe@liu.edu.lb

***Abstract* - As any technology, medical equipment provides benefits to patients, but they also present significant risks that can affect and threaten patient safety. In healthcare organizations, clinical engineering departments play a big role in maintaining the safety and reliability of medical equipment. In order to mitigate failures of such equipment and control risks, a proper Medical Equipment Management Program (MEMP) should be established. The purpose of this paper is to forecast risks by using Failure Mode and Effect Analysis (FMEA) method and apply it on Monte Carlo simulation which adds risks analysis to Excel® by @RISK tool. The data of some medical devices is extracted from a hospital's maintenance management system and are identified according to their likelihood, severity, and difficulty of detection. However, the results of this mathematical simulation are integrated in a probability distribution function that enable us to identify medical equipment risks that affect patients, staff, and the work environment and reduce them by providing contingency plans, policies, strategies, and other tactics [1].**

Keywords - medical equipment; risk management; FMEA; Monte Carlo simulation; HIQMA.

I. INTRODUCTION

As medical technology becomes more complicated, a MEMP must be applied to ensure that medical devices operate according to safety, accuracy, reliability, and performance criteria. Maintenance is one of the most important processes to improve safety, decrease the risk of equipment failure, and minimize the unplanned downtime [2]. However, the money spent on maintenance and failure of equipment is rapidly increasing because of the development of many types of complex medical equipment, the stringent environment they are operating under it, and the lack of proper management.

The management program includes a risk management process, which comprises the identification, assessment, and prioritization of risks (defined in ISO 31000 as the effect of uncertainty on objectives) followed by coordinated and economical application of resources to minimize, monitor, and control the probability and/or impact of unfortunate events [3]. The causes of the risks are identified and relevant changes in the system are made accordingly in order to reduce the probability of the error occurring in the future thus reducing harm to patients and providing a safer patient care experience.

Most healthcare organizations follow the manufacturer's recommendations concerning the maintenance program [4]. Campbell and Jardine [5] defined the maintenance excellence as the balance of performance, risk, resource inputs and cost to reach to an optimal solution. In the last decade, maintenance techniques have been notably improving, but most of the healthcare organizations do not profit from the maintenance excellence that Campbell and Jardine established. Moreover, some devices that are similar in their function and design have manufacturer-recommended intervals that vary by one or two factors thus leading to financial and time loss. In addition, excessive maintenance can have the same impact as an insufficient level of maintenance; moderation should be the rule.

The status of research on maintenance of medical devices is presented in different models. Fennigkoh and Smith [6] model classified equipment according to three parameters: function, physical risks, and maintenance requirements. It was known later as risk-based inclusion criteria that allowed clinical engineering professionals to apply maintenance on limited parts of medical devices.

Ridgway, in the beginning, noted that Preventive Maintenance (PM) is an important factor in terms of reliability, but later on, he indicated that PM does not prevent failure for all equipment and it is not the ideal solution. However, Ridgway provided methods for equipment management such as Reliability Centered Maintenance (RCM). This latter is a corporate-level maintenance strategy that is implemented in any healthcare organization to optimize the maintenance program. Endrenyi [7] indicated that RCM selects the critical component in the equipment and starts a maintenance management to correct the failure. Further on, he recognized that RCM is good for indicating the budget and for comparing policies, but it cannot help in achieving real optimization.

According to Hall [8], the two keys of RCM are having a good maintenance history of the medical equipment and the age of the equipment. Further he indicated that RCM is applicable for younger equipment. However, to balance between preventive and corrective maintenance, Condition Based Maintenance (CBM) is presented to observe and forecast real time status of machines [8]. CBM is performed when some indicators show that the equipment will fail.

Taghipour et al. [9] presented a multi-criteria decision-making model to prioritize medical devices according to their criticality. Furthermore, in terms of prioritization, Jamshidi et al. [10] developed a fuzzy healthcare failure modes and effect analysis (HFMEA). HFMEA is a systematic method that identifies and prevents equipment problems before they occur by ensuring a safe and clinically desirable outcome [11].

To minimize risk and optimize the cost-effectiveness of medical equipment, a maintenance model is suggested by Khalaf et al. [12]. They evaluated both elements and the results showed poor performance concerning cost and risk management. Therefore, Khalaf et al. [13] developed a new model in order to be used in Palestinian hospitals, which is a mathematical model that uses a mixed integer-based approach for maintenance operations schedules for medical equipment. They also proposed a greedy algorithm for an initial solution for the model. In addition, some data extracted from maintenance history of infusion pumps and ventilators were used in a global model that measures the probability of equipment being available and they were analyzed using Matlab. However, this model was validated by developing a model that measures the survival of equipment as function of maintenance and age of equipment using survival analysis approach.

The studies reported above proposed models that share a common theme; different risks are calculated using a single measure that is defined and used to lead safety, performance inspections, and preventive maintenance activities. These models are simple to use and effective in reducing general risks yet they lack the ability to identify specific risks. They are far from achieving optimal risk minimization. Also, research into comprehensive frameworks for prioritizing critical medical devices or outsourcing of medical device maintenance is still in its infancy. Researchers should apply new risk-based maintenance models including different new uncertainties to replace the traditional empirical models.

In our model, a Complete Risk and Decision Analysis Toolkit from Palisade: "The DecisionTools Suite" is used. It is an integrated set of programs for risk analysis and decision-making under uncertainty that runs in Microsoft Excel®. The main tool that was used is @RISK, which adds risk analysis to Excel® using Monte Carlo simulation. The Monte Carlo simulation is a technique used to understand the impact of risk and uncertainty in financial, project management, cost, and other models which is to identify risks related to medical equipment [14]. FMEA method was also used to prevent failure of equipment. Data related to maintenance and failures of equipment were obtained from a Lebanese hospital to apply them in our model in order to verify its functionality and applicability.

The proposed methodology is presented in Section II. The implementation process is presented in Section III. This latter, includes collecting data, and integrating FMEA method using Monte Carlo simulation. This is followed by results and discussion in Section IV. Finally, a conclusion and our further expectations are presented in Section V.

II. Methodology

Medical devices are used in healthcare organizations to support patient care in terms of health and safety. Currently, modern medical devices are complex and operate under severe conditions because of the rapid development of equipment. The current strategies in hospitals have difficulties in identifying risks and applying optimal risk reduction activities because they lack proper management processes. Therefore, a well-operated management process can enhance the function of medical devices in healthcare organizations.

The proposed model is meant to identify and assess risks of medical equipment according to mathematical approach using different parameters. It starts with collecting data concerning medical devices from a Lebanese Hospital. The needed numbers such as the likelihood, detectability, and impact of medical equipment failure are then extracted and analyzed.

There are several methods to calculate the risk value, yet FMEA method is used as the preferred choice in the current model. FMEA is selected among other methods because it contributes to improved designs for products and processes, to cost savings, and to the development of control plans, testing requirements, optimum maintenance plans, reliability growth analysis and related activities [15]. The FMEA procedure starts with determining the ways in which the input can go wrong, and then determining effects for each failure mode. After that, it identifies potential causes for each mode and list current controls for each cause. Consequently, risk priority number can be determined and contingency plans and actions should be set accordingly.

After applying the FMEA method, it will be integrated in Monte Carlo simulation tool that includes @Risk toolkit. @Risk adds risk analysis to Excel® using Monte Carlo simulation. Then the simulation will be performed and the results will be assessed to draw a conclusion.

The methodology is depicted in the following flowchart.

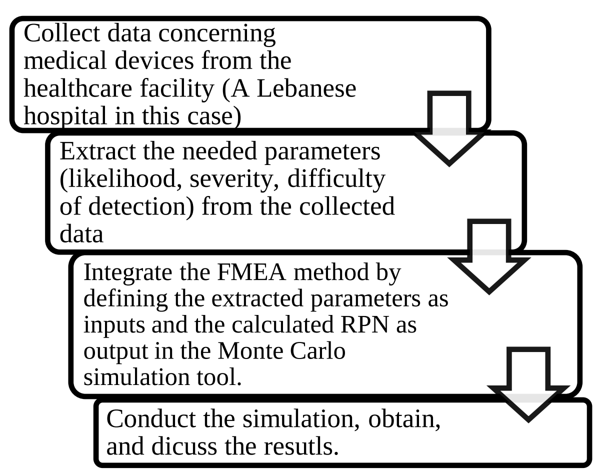


Figure 1. The proposed methodology

Fig. 1 summarizes the required steps to accomplish our evaluation. Such assessment requires some parameters and equations. So, the derivations of all those relations are explained in the subsequent sections.

III. Implementation Process

The implementation process includes three main steps that are “Collecting and Extracting Data”, “Integrating FMEA Method in Monte Carlo Simulation Tool”, and “Simulation and Results” to be accomplished.

A. *Collecting and Extracting Data*

First, to apply the FMEA method, specific data concerning medical devices are collected.

Likelihood of the medical device in this case is the probability of failure of the machine. Fig. 2 shows the number of repeated failures per year with respect to medical devices. These numbers are then converted to a scale of 1-10 as shown in Table I using the following conversion:

Number of repeated failures*(10/ Highest number of repeated failures)

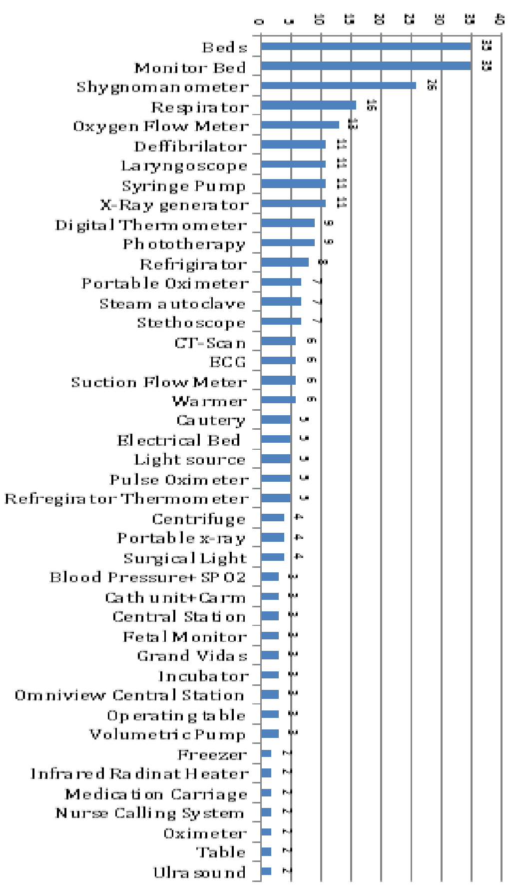


Figure 2. Number of repeated failure.

The scores of likelihood of medical devices failures are assigned according to the following criteria [16]:
{1, 2}: Improbable, manifestations of the hazard are very unlikely
{3, 4}: Remote, manifestations of the hazard are possible but not likely
{5, 6}: Occasional, some manifestations of the hazard are likely to occur
{7, 8}: Probable, hazard will be experienced
{9, 10}: Frequent, hazard likely to occur

Severity of medical device is defined as the extent to which the defect of equipment can affect patients. The scores of severity are assigned according to the following criteria [16]:
{1, 2}: Negligible, no significant risk of injury
{3, 4}: Minor, potential for minor injury
{5, 6}: Moderate, potential for minor injury
{7, 8}: Critical, potential for severe injury
{9, 10}: Catastrophic, likely to result in death

Detection is the ability of the current control scheme to detect and then prevent a hazard from occurring. The scores of detection are assigned according to the following criteria [11]:
{1, 2}: Almost certain (detection probability <100%), potential hazard will almost certainly be detected
{3, 4}: High (detection probability <80%), high chance that potential hazard will be detected
{5,6}: Moderate (detection probability <50%), moderate chance that potential hazard will be detected
{7,8}: Low (detection probability <25%), low chance that potential hazard will be detected
{9, 10}: Remote (detection probability <10%), very remote chance that potential hazard will be detected

Table I. Extracted Parameters.

Equipment	Likelihood	Severity	Difficulty of Detection
Beds	10.00	6	1
Sphygmomanometer	7.43	5	1
Defibrillator	3.14	10	4
Ultrasound	0.57	3	3
Pulse Oximeter	1.43	7	2
Syringe Pump	3.14	10	7

All values of likelihood, severity, and difficulty of detection of equipment are given by the hospital. Table I shows the scores of likelihood, severity, and difficulty of detection for six medical devices on a scale of 1-10.

B. *Integrating FMEA Method in Monte Carlo Simulation Tool*

The parameters extracted from the collected data will be employed in a systematic technique called FMEA.

FMEA is one of the first highly structured, systematic techniques for failure analysis. It was developed by reliability engineers in the late 1940's to study problems that might arise from malfunctions of military systems [17]. It is a step-by-step systematic approach for identifying all possible failures in a design, a manufacturing or assembly process.

Failures are prioritized according to how severe their consequences are, how likely they may occur and how difficult to detect them. The main purpose of the FMEA is to take preliminary actions to reduce failures, starting with the highest-priority ones [18].

Risk Priority Number (RPN) is a measure used when assessing risk to help identify critical failure modes. The RPN values range from 1 (absolute best) to 1000 (absolute worst). It is the product of three ratings on a scale of 10 (likelihood of occurrence, severity of impact, and difficulty of detection):

$$RPN = Likelihood * Severity * Difficulty\ of\ Detection$$

Table II illustrates the extracted parameters and the calculated RPN for each equipment.

Equipment	Likelihood	Severity	Difficulty of Detection	RPN
Beds	10.00	6	1	60.00
Sphygmomanometer	7.43	5	1	37.15
Defibrillator	3.14	10	4	125.60
Ultrasound	0.57	3	3	5.13
Pulse Oximeter	1.43	7	2	20.02
Syringe Pump	3.14	10	7	219.80

Table II. Calculated RPN.

After calculating the risk priority numbers, the model is now ready to be integrated in the @Risk simulation tool.

The first step is to insert Table II in an Excel® sheet and define inputs (likelihood, severity and difficulty of detection) as normal distributions. Usually, high standard deviation is selected in situations where resources are limited or gathering real data would be too expensive or impractical. In this situation, the data is extracted from a real hospital management system, hence a very small standard deviation is selected (0.1), as depicted in Fig. 3:

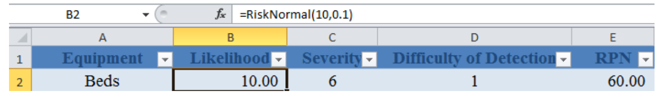


Figure 3. Definition of inputs as normal distributions.

RPN is the output in our model; Fig. 4 illustrates how RPN is defined as an output in the Monte-Carlo simulation tool "RiskOutput("RPN")".

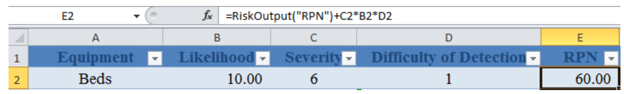


Figure 4. Adding @Risk output.

C. *Simulation and Results*

@RISK monitors a set of convergence statistics on each output distribution during a simulation. During monitoring, @RISK calculates these statistics for each output at selected intervals (such as: every 1000 iterations) throughout the simulation.

As more iterations run, the amount of change in the statistics becomes less and less until they reach the Convergence Tolerance [19].

Convergence tolerance specifies the tolerance allowed for the statistic being tested. For example, the current applied settings specify that the estimated mean of each output is simulated within 3% of its actual value [19].

In our model in Fig. 5, we will be performing 5000 iterations in one simulation.

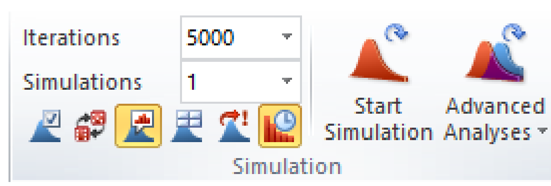


Figure 5. Changing the number of iterations and starting simulation.

At the end of the simulation, the results are integrated in a probability distribution function. A probability distribution is a statistical function that describes all the possible values and likelihoods that a random variable can take within a given range [20]. This range will be between the minimum and maximum statistically possible values, but where the possible value is likely to be plotted on the probability distribution depends on a number of factors, including the distributions mean and standard deviation.

Fig. 6 illustrates one example of the results obtained; the risk priority number of hospital Beds (60) is centered between 49.98 and 70.09 for 90% of the probability distribution. The x-axis represents the possible risk priority numbers and the y-axis represents the probability of occurrence for each probable RPN incrementing by 0.02 on a scale of 0.1.

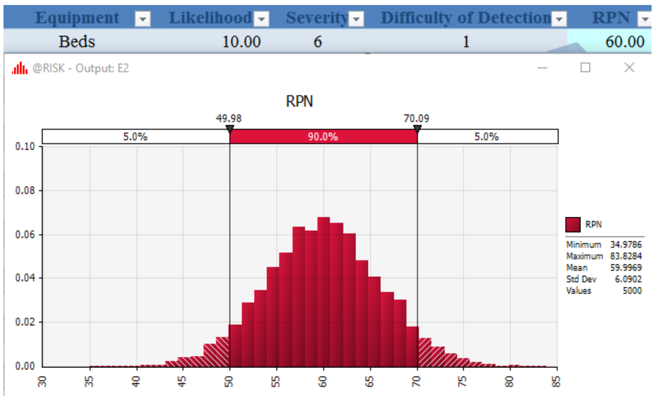


Figure 6. Results after simulation.

IV. Results and Discussion

The result of the Monte Carlo Simulation via @RISK is a probability distribution. Figs. 6, 7, 8, 9, 10, and 11 show the probability density for the chosen examples: beds, defibrillator, ultrasound, syringe pump, sphygmomanometer, and pulse oximeter.

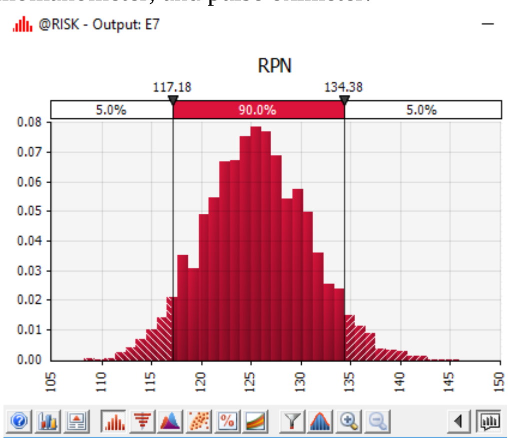


Figure 7. Probability distribution for defibrillator.

Fig. 7 shows the risk priority number of defibrillator (125.60) is centered between 117.18 and 134.38 for 90% of the probability distribution.

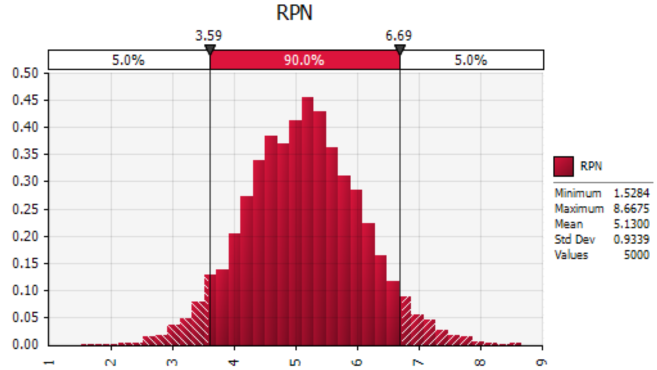


Figure 8. Probability distribution for ultrasound.

Fig.8 shows the risk priority number of ultrasound (5.13) is centered between 3.59 and 6.69 for 90% of the probability distribution.

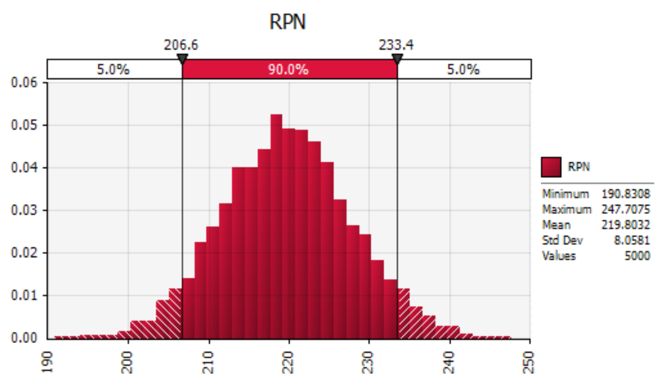


Figure 9. Probability distribution for syringe pump.

Fig.9 shows the risk priority number of syringe pump (219.80) is centered between 206.6 and 233.4 for 90% of the probability distribution.

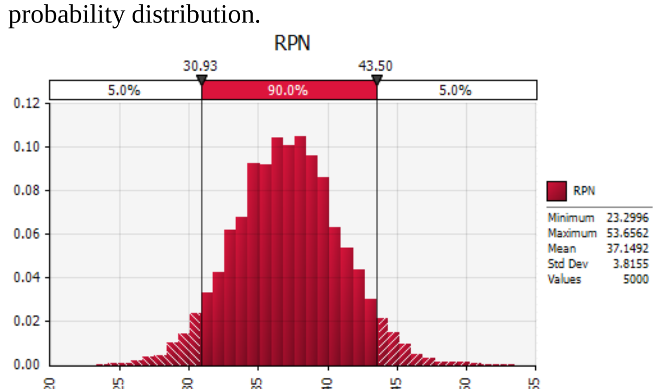


Figure 10. Probability distribution for sphygmomanometer.

Fig.10 shows the risk priority number of sphygmomanometer (37.15) is centered between 30.93 and 43.50 for 90% of the probability distribution.

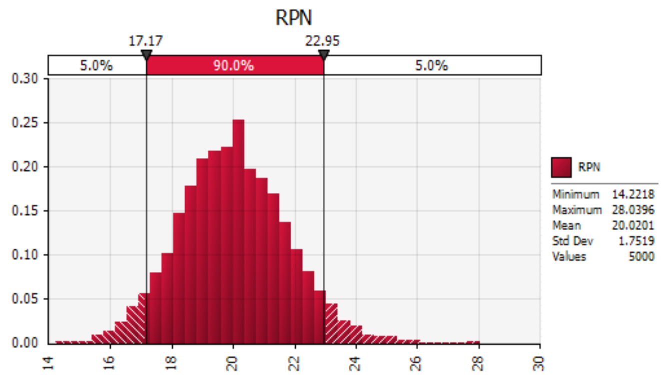


Figure 11. Probability distribution for pulse oximeter.

Fig.11 shows the risk priority number of pulse oximeter (20.02) is centered between 17.17 and 22.95 for 90% of the probability distribution.

	Minimum	Maximum	Mean
Beds	34.97	83.82	59.99
Sphygmomanometer	23.29	53.66	37.15
Defibrillator	108.23	145.19	125.6
Ultrasound	1.53	8.67	5.13
Pulse Oximeter	14.22	28.04	20.02
Syringe Pump	190.83	247.70	219.80

Table III. Summary of the Results.

The results presented in Figs. 6, 7, 8, 9, 10, and 11 and Table III is interpretable as follows:

1. The mean figure for RPN will be 60 for beds, 37.15 for sphygmomanometer, 125.6 for defibrillator, 5.13 for ultrasound, 20.02 for pulse oximeter and 219.80 for syringe pump. That means, the simulated result will be equal to the original calculated RPN.

2. The minimum figure for RPN will be 34.97 for beds, 23.29 for sphygmomanometer, 108.23 for defibrillator, 1.53 for ultrasound, 14.22 for pulse oximeter and 190.83 for syringe pump. That means, the minimum probability will be lower than the calculated RPN by 25.03 for beds, 13.86 for sphygmomanometer, 17.37 for defibrillator, 3.6 for ultrasound, 5.8 for pulse oximeter and 28.97 for syringe pump. But theses figure are the bottom lines and will only be achieved if all negative circumstances would occur. Hence, with a probability of 5 %, the figure for RPN will fall low to 34.97, 23.29, 108.23, 1.53, 14.22, and 190.83. In other words, with a probability of 95 % the RPN will not fall below these numbers.

3. The maximum figure for RPN will be 83.82 for beds 53.66 for sphygmomanometer, 145.19 for defibrillator, 8.67 for ultrasound, 28.04 for pulse oximeter and 247.70 for syringe pump. That means, the maximum probability will be higher than the calculated RPN by 23.22 for beds, 16.51 for sphygmomanometer, 19.59 for defibrillator, 3.54 for ultrasound, 8.02 for pulse oximeter and 27.9 for syringe pump. But these figures are the upper limits and will only be achieved if all positive circumstances would occur. Hence, with a probability of 95 %, the figure for RPN will not exceed 83.82, 53.66, 145.19, 8.67, 28.04, and 247.70. In other words, with a probability of 5% the RPN will exceed these numbers.

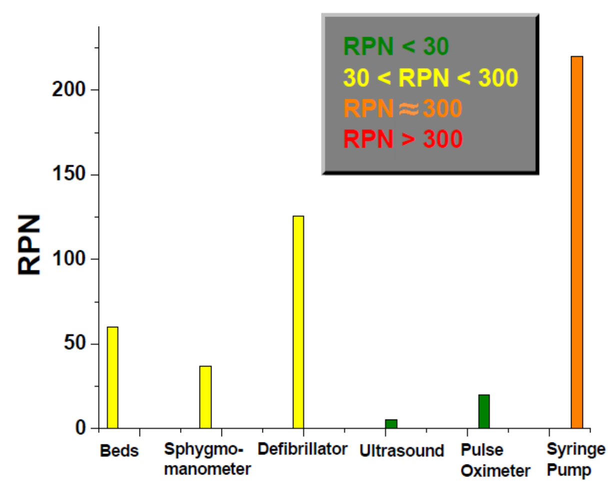


Figure 12. Histogram of RPNs

Fig. 12 shows a histogram illustrating each device's RPN. This histogram helps us track the severity of risks on each device in order to solve the problem before happening. Each of the RPN scores will fall under one of the categories, for which different colors have been used. Here are some details on each of the categories:

High Risk: It represents the red color which is the most dangerous category. Example: Syringe pump: almost in the high risk category (presented in orange).

Medium Risk: The yellow category has less priority than the one before but also plans and decisions must be set to handle those risks. Example: Beds, Sphygmomanometer, and Defibrillator.

Low Risk: The last category that represents the green color has the lowest priority where risks can be monitored minimally, and do not cause serious problems. Example: Ultrasound, Pulse oximeter.

After analyzing the results, some recommendations could be set to reduce risks such as having alternative or redundant devices in the healthcare facility, pay special attention to the to the life span of the equipment and its working hours when purchasing used devices, and to have a well operated maintenance program. Moreover, hospitals must be kept financially healthy while achieving financial-related risk management goals for healthcare organizations by reducing the malpractice claims and the number of failures.

An additional evaluation is possible to show where individual risk has a main influence of the final risk priority number. Figs. 12, 13, 14, 15, 16, and 17 show the results of those evaluations as regression coefficients. This indicates that difficulty of detection has a huge influence of the RPN of beds and the likelihood has the higher influence on the RPN of the defibrillator. Therefore, these risk factors have to be monitored very carefully within an effective healthcare management system.

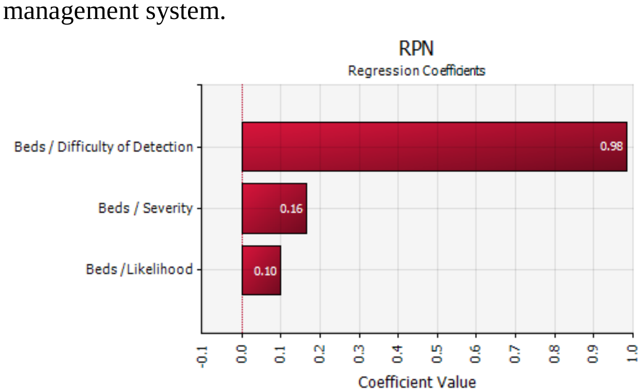


Figure 13. Regression coefficients for beds.

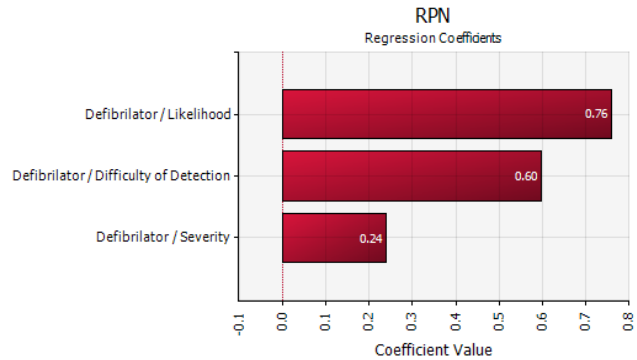


Figure 14. Regression coefficients for defibrillator.

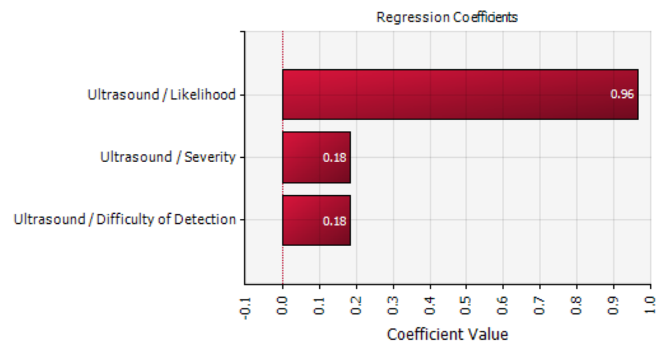


Figure 15. Regression coefficients for ultrasound.

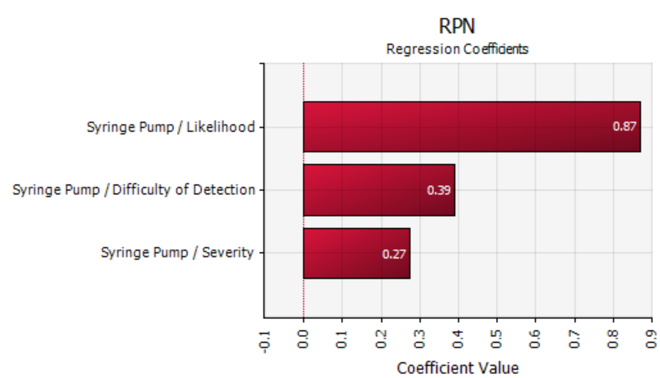


Figure 16. Regression coefficients for syringe pump.

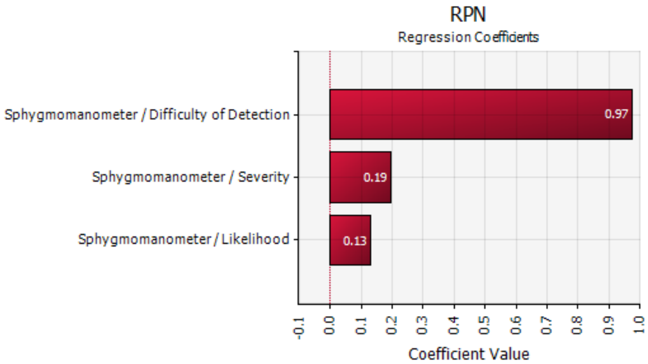


Figure 17. Regression coefficients for sphygmomanometer.

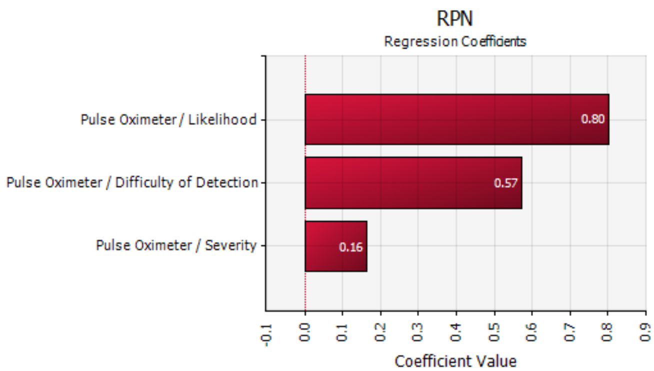


Figure 18. Regression coefficients for pulse oximeter.

However, this model could be integrated in the HIQMA (Hospital Institution Quality Management) system. HIQMA was deployed for the first time in Lebanon in early 2011 to enhance medical and healthcare services ensuring quality [21]. The principle objective is to guarantee patients' safety through viable and effective quality management in system includes scalability and customizability traits. The framework properties incorporate a few applications, beginning with the individual beneficiary organizations, proceeding with the reformed professional training and advisory services concepts, and ending with the created administration rules [21].

HIQMA is a centralized management system that provides a gateway to critical quality information and facilitates quality performance improvement through requirement tracking, notifications and real-time management reporting [21]. This system increases marketability, customer satisfaction and service; it also saves time, money and resources. Moreover, it improves internal communication and operational performance and provides better management control [22].

Biomedical engineers and technicians using HIQMA are providing the system with important dated information included in their reports about each equipment failure, thus providing enough data to calculate its probability of occurrence. Also, they are mentioning the severity and consequences of each occurred failure. The only missing information in the reports is the difficulty of detection; it can be collected from their experience in a questionnaire included in each technical report on a scale from 1-10.

Finally, a risk severity matrix is employed to raise awareness and increase visibility of risks so that sound decisions on certain risks can be made. The risk matrix is shown in Fig. 18. Once the risks have been placed in the cells of the matrix that corresponds to the appropriate likelihood, severity and difficulty of detection, it becomes visibly clear as to which risks must be managed at what priority.

Each of the risks will fall under one of the categories, for which different colors have been used. Table IV represents the letter of each device and under what color it falls. Here are some details on each of the categories:

High: The risks that fall in the cells colored in red are the risks that are most critical and that must be addressed on a high priority basis. Example: ‘X’ Defibrillator, ‘Y’ Syringe Pump.

Medium: If a risk falls in one of the yellow cells, it is best to take some reasonable steps and develop risk management strategies in time, even though there is no hurry to have such risks dealt with early. Example: ‘C’ Sphygmomanometer, ‘I’ Beds and ‘O’ Pulse Oximeter.

Low Risk: The risks that fall in the green cells can be minimally monitored as they usually do not pose any significant problem. However, if some reasonable steps can help in fighting these risks, such steps should be taken to improve overall performance Example: ‘A’ Ultrasound Machine.

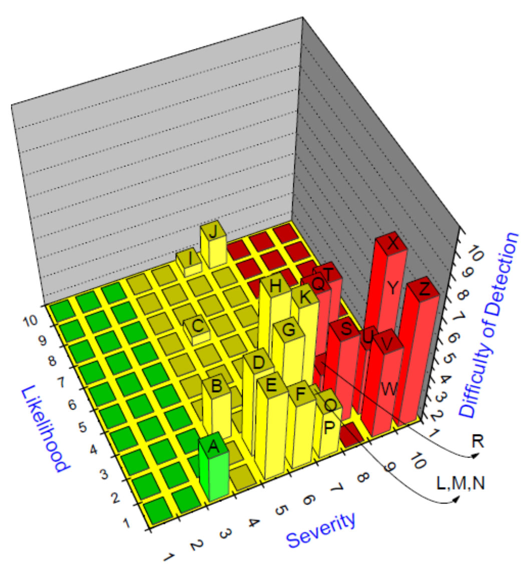


Figure 19. Risk severity matrix.

Table IV. Legend for the risk severity matrix.

Ultrasound	A	green
X-Ray Generator	B	yellow
Sphygmomanometer	C	yellow
CT-Scan	D	yellow
Portable X-Ray	E	yellow
Centrifuge	F	yellow
Warmer	G	yellow
Digital Thermometer	H	yellow
Beds	I	yellow
Monitor Bed	J	yellow
Phototherapy	K	yellow
Portable Oximeter	L	yellow
Stethoscope	M	yellow
ECG	N	yellow
Pulse Oximeter	O	yellow
Blood Pressure + SPO2	P	yellow
Oxygen Flow Meter	Q	red
Laryngoscope	R	red
Suction Flow Meter	S	red
Respirator	T	red
Cautery	U	red
Cath Unit	V	red
Incubator	W	red
Defibrillator	X	red
Syringe Pump	Y	red
Volumetric Pump	Z	red

After analyzing the results, some recommendations could be set to reduce risks such as having alternative or redundant devices in the healthcare facility, pay special attention to the to the life span of the equipment and its working hours when purchasing used devices, and to have a well operated maintenance program.

V. Conclusion and Future Work

The rapid evolution of medical equipment had a huge impact on the improvement and progress of medical services. Accordingly, medical devices are expected to operate under safety, accuracy, and reliability criteria to ensure a protected and efficient environment for patients, staff, and the surrounding work environment. As such, this research work provided a new methodology for identifying and assessing risks based on a mathematical approach and not only empirical ones. This method results in a more precise scheme that would most likely reduce the risks resulting from medical equipment and further provide a proper management in healthcare organizations. Moreover, this model can be integrated in healthcare facilities to identify and forecast risks according to risk distribution of Monte Carlo simulation and risk severity matrix that classifies and prioritizes medical devices risks.

This proposed assessment maybe be further enhanced to achieve risk response development, and risk response control of medical equipment by developing a complete tool that can be used in the medical equipment industry across the world. There should also be a further research in the field of optimal outsourcing of medical devices. Thus, manufacturers, organizations, and clinical engineering departments can use this tool in planning for maintenance and for the development of medical equipment. Also, it can be deployed as monitoring system in service at healthcare facilities where it can provide real time data on the risks of operating medical equipment.

VI. References

[1] S. Ismail, H. Nehme, M. Hajj-Hassan, H. Bassam and H. Khachfe, "A Monte Carlo Simulation Based-Approach for Medical Equipment Risks Forecasting," in *Global Health*, 2016.

[2] K. Wilson, K. Ison and S. Tabakov, Medical Equipment Management, CRC Press, 2013.

[3] D. Hubbard, The Failure of Risk Management: Why It's Broken and How to Fix It, John Wiley & Sons, 2009.

[4] A. Jamshidi, S. Abbasgholizadeh-Rahimi and D. Ait-Kadi, "Medical Devices Inspection and Maintenance; A Literature Review," in *Industrial and Systems Engineering Research*, Canada, 2014.

[5] J. Campbell and A. Jardine, Maintenance Excellence:Optimizing Equipment Life-Cycle Decisions, CRC Press, Feb. 13, 2001.

[6] L. Fennigkoh and B. Smith, Clinical equipment management, vol. 2, JCAHO PTSM Series, 1989, pp. 5-14.

[7] J. Endrenyi, "Aging maintenance and reliability," *IEEE Power & Energy Magazine*, pp. 59-67, May 2006.

[8] A. Hall, Risk-based approach to device management has multiple pathways, vol. 40, Biomedical Instrumentation & Technology,, 2006, pp. 225-226.

[9] S. Taghipour, D. Banjevic and A. Jardine, "Prioritization of Medical Equipment for Maintenance Decisions," *Journal of Operational Research Society*, vol. 9, no. 1666-1687, p. 62, 2010.

[10] A. Jamshidi, S. Abbasgholizadeh-Rahimi, D. Ait-kadi and A. Ruiz, "A risk-based Maintenance Strategy using Fuzzy HFMEA for prioritization of Critical

Medical Equipment," in *Proceedings of the 1st Annual World Conference of the Society for Industrial and Systems Engineering*, Washington, USA, 2012.

[11 Maintenance Assistant™, [Online]. Available: https://www.maintenanceassistant.com/reliability-centered-maintenance/. [Accessed October 2015].

[12] A. Khalaf, Maintenance model for minimizing risk and optimizing cost-effectiveness of medical equipment in Palestine, vol. 14, JClinEng, 2004, pp. 3649-3653.

[13] A. Khalaf, K. Djouani, Y. Hamam and Y. Alayli, "Evidence-based mathematical maintenance model for medical equipment," in *International Conference on Electronic Devices, systems and Applications*, 2010.

[14] Risk Analysis and Simulation, New York: Palisade Corporation, September, 2010.

[15] L. Lipol and J. Haq, "Risk Analysis Method: FMEA/FMECA in the Organizations," *International Journal of Basic & Applied Sciences*, vol. 11, October 2011.

[16] llloyd. [Online]. Available: http://www.mddionline.com/article/risk-analysis-beyond-probability-and-severity. [Accessed November 2015].

[17] U.S. Department of Defense, Procedures for Performing a Failure Mode, Effects and Criticality Analysis, 1949.

[18] [Online]. Available: http://asq.org/learn-about-quality/process-analysis-tools/overview/fmea.html. [Accessed February 2016].

[19] Palisade Corporation, *Guide to Using @Risk: Risk Analysis and Simulation,add-In for Microsoft® Excel*, 5.7 ed., 2010.

[20] [Online]. Available: http://www.investopedia.com/terms/p/probabilitydistribution.asp. [Accessed March 2016].

[21] B. Hussein, H. Bazzi, W. Hassan, A. Hage-Diab, A. Haj-Ali and H. Khachfe, "Software Implementation of a Healthcare Quality Management System Based on ISO9000 Standards," *SM Journal of Biomedical Engineering*, 2016.

[22] [Online]. Available: http://www.bsigroup.com/en-GB/iso-9001-quality-management/. [Accessed February 2016].

Digitizing Health Care in Collaboration Between Nursing and Engineering

Two cases of strategic learning and implementation of robots in the homes of elderly people

Britt Östlund
Department of Biomedical Engineering and Health Systems
Royal Institute of Technology, KTH
Stockholm, Sweden
brittost@kth.se

Gunilla Björling, Sara Stridh, Madeleine Sahlström, Janet Mattsson
Swedish Red Cross University College
Stockholm, Sweden

***Abstract* - Digitization exceeds the limits of healthcare meetings, which gives renewed relevance to examine the collaboration between engineers and nurses. Caring for people is no longer just something going on in the hospital but at new arenas at home and in the middle of people's everyday lives. In caring situations nurse's responsibility is, unlike the physician, to make observations and to follow in detail the patient's caring needs, and where engineers provide technological devices to support and monitor the course of the disease. When digitizing the caring situation person-centered care gets a new meaning. For engineers the understanding of how technology is contextualized and domesticated becomes even more important to make applications and systems work outside laboratories. This paper presents two cases of interaction between engineers and nurses aimed at improving the implementation of robots and sensors in elderly people´s homes; and learning how to improve patient safety in hospitals. The result shows that conflicting epistemologies, differences in professional languages and lack of joint learning opportunities are factors that create obstacles for interactions. The conclusions reject the idea of linear innovation processes and show that successful ccollaboration take more than just adding two and two together. Especially digitization is breaking up traditional barriers and hierarchies. For nurses to be proactive requires knowledge about technological developments and the ability to participate in design and innovation processes. For engineers a more thorough understanding of caring situations and users will contribute to a more reliable provision of digital solutions and point at new ideas leading up to innovations. The main output of the paper is that it is deepening the understanding of what factors leading to successful collaborations between nursing and engineering and what are the missing links.**

Keywords-digitization; caring; nursing; engineering

I. INTRODUCTION

A proposal for a new research subject including nurses and engineers was presented at the Ninth International Conference on eHealth, Telemedicine and Social Medicine, eTelemed 2017 [1]. The background is the fact that today´s health care systems face a number of challenges related to technological developments. The Global Commission on Education of Health Professionals for the 21st Century describes a mismatch between the care offered and people's demands and needs; lack of cooperation; discontinuous care chains; tenacious hierarchies, and not least, a focus on technology founded on flawed understanding of the context in which the technology is used [2]. The gender system that locks structures what perceived as male and female work is deemed particularly difficult to change. Significant is also the lack of good examples of how to meet these types of challenges. Other publications confirm these results, reporting on a lack of accuracy in technological support for health care, not least the confidence in individual solutions. There are reports on an imbalance between the success factors in health care - increased life expectancy and better treatments - and the expectations of what you want done and what to expect [3] [4]. At the same time, the belief in technology as the ultimate solution still prevails. The Swedish Society for Nursing identifies a lack of knowledge about the impact of technology on nursing interventions. To meet this, they emphasize individual health as an important aspect to investigate and call for possibilities to actively participate in design, implementation and assessment of new technology [5] [6]. Others stress the economic value of returning to patient centred strategies [7]. To mention one example, creating digital health care records and digital information to citizens are among the most difficult tasks to accomplish despite the wide access to Internet and digital infrastructures. Google declared a few years ago that their investment in Google Health was one of their biggest failures due to its complexity [8]. Instead they are now developing databases focusing on health and aging sciences.

A. The caring situation in focus

A question that has been discussed for a long time to make development and implementation work is the collaboration between engineers and health care professionals. Ever since C.P. Snow presented his thesis on the two cultures in 1959 stating that "the intellectual life of the whole of western society" was split into the two cultures — sciences and the humanities — creating a major obstacle for solving the world's problems [9], the need for multidisciplinary collaboration has grown in importance. Today there are a number of examples of international multidisciplinary collaboration, e.g., “the Bio design Fellowship program” at Stanford University since 2003 which was also implemented in Sweden as The Centre for Technology in Medicine and Health (CTMH), the Canadian Newfoundland and Labrador Support program, or the Erasmus MC program in Netherlands taking on a patient perspective, just to mention a few. These types of commitments are key elements in promoting interdisciplinary research and development in medicine focusing problem-solving in general between stakeholders, health care professionals and patients.

However, it is rare to find evidence for how to create successful collaborations focusing the caring situation including nurses and explanations to why this is difficult to achieve. Biomedical engineering and clinical trials are for the most part related to doctors collaborating with engineers. Lately the need for nurses to be involved is discussed [10]. especially when it comes to teamwork [11] [12]. Inter-professional collaboration during education is pointed out as a good opportunity to overcome professional barriers [13]. and some experiences published [14]. This is especially asked for in times of increasing specialization [15]. Beyond this, interventions into caring situations and nursing are for the most part about studying them, interviewing them about their work environment [16] [17] [18] [19], musculoskeletal disorders or other physiological problems in working life [20] [21], job demands [22], exposure for abuse or other risks [23] [24] [25] [26] [27], workload [28] or to prevent or plan for interventions [29] [30] [31]. Interventions with a participative approach include e-training programs [32], ergonomic preventative programs [33] [34] or integrated care [35] or are discussing success factors for interventions in nursing in general [36]. However, there are no conclusions about what limits multidisciplinary collaborations or what makes it possible.

When turning the health care meeting, or caring situation, as it will hereinafter be called, into the research object and focus for this paper we do not include any health care professionals. The caring situation is defined as the situation where nurses and biomedical engineers are both interfering with caring for the patient. Unlike the physician who is responsible for making diagnosis, it is the nurse who has the responsibility for the caring situation. Nurses are making observations following in detail the patient's caring needs and providing relevant and sufficient care while engineers provide technological support for monitoring the course of the disease. In a critical situation where technology fails or the patient status changes in an unexpected way, engineers immediately start to search for technological problems while nurses immediately direct his or her concentration on the patient. This is good and well and according to their training and competencies, but at the same time they interfere with the same patient and in the same situation, and not least in connection with digitization and implementation in new arenas. One can assume that their interaction and communication is a vital factor that influences the outcome of critical situations and the patient’s wellness. Furthermore, digitization provides us with new challenges and opportunities to examine the collaboration.

B. Engineering and nursing when digitizing

The initiative presented in this paper takes place in the era of digitization and digitizing health care is the overall context in which this is under development, in fact one of the most important advances, compared to when the healthcare system developed on a scientific basis [2]. At that time, modern science was integrated in medical training leading up to reforms and knowledge that doubled the life expectancy during the 20th century. Today digitization encompasses bio-medical engineering used by health care professionals in hospitals, and home health care systems, as well as digital tools used by both health care professionals, citizens at home and in mobile settings.

Engineering and nursing, including caring sciences, have a long history together in developing modern health care. While engineers have provided tools and instruments in accordance with technological advancements, nurses have had the role as users of these applications in patient contexts and have been responsible for the safety around the patient. In this context engineering includes technical research across the borders of engineering and medicine important to medical applications and health care in its widest meaning encompassing research on cellular and molecular level to complex systems and materials and energy. Traditionally, doctors at hospitals are pointed out as the target group for medical technologies. With digitization breaking up hierarchies there are reasons to point out other health care professions beyond the role of the doctors, nurses having a central role close to patients.

In the Nordic countries, nursing as a research subject, have for thirty years, developed in parallel with “caring sciences”. Both have been expansive including caring informatics and caring theory. Nursing, which is the broad international field, has its focus on guiding nurses in practice such as routines and regulations and patient safety. Caring sciences, published in Journals, such as the Scandinavian JN for Caring Sciences among other JN, originate from phenomenology and the interest to understand principles for utilization [37] [38] [39]. Caring science is today related to person-centred care, self-sufficiency and independence. Especially since the core of caring has been revealed as central, holistic, individualized and at the same time providing expert physical care combined with fulfilling emotional needs in an adaptive environment [40].

Digitization takes place in parallel with another characteristic of our time - demographic developments – in fact the most contemporary social change of our time taken place today in most parts of the world. Drastic reductions of child mortality and increased living standards in developing countries have increased the average life expectancy. Children and old populations constitute the groups that will increase the most until 2050, hence being the main patient target groups for the coming decades both locally and globally. It is estimated that within a decade, the majority of the world’s population will have accesses to virtually all the world's information in a machine that fits in its own palm. These profound changes occurring within one generation naturally has a great impact on digital media, robots and sensors creating new opportunities to practice disaster response and care, in dangerous situations and in the monitoring of health.

In this paper digitization is understood in its historical context being an extension of the use of IT, converting even more information into digital form. The telephone system used by broad groups of citizens is one example of an everyday technology with multiple uses now being digitized. A long used technology is the safety alarm that has been in operation in home care since the 1970s and now being digitized. There are good reasons to assume that previous experiences of use and technological changes affect individuals approach to new technologies. Following the classic theory of Everett Rogers, adoption is a process that deals with the uncertainty in deciding about a new alternative to those previously in existence [41]. From this theory the widespread use of the safety alarms can probably be explained by the long experience of using telephone applications. Second, digitization occurs in everyday life, not in limited and controlled environments, since exchanges between people are mediated digitally to an increasing degree. This provides more complex challenges since most of medical technologies traditionally are developed and tested in laboratories away from real life settings. To understand why digital applications and systems do work or do not work in people´s everyday life requires a subtler and theoretical understanding of contextualization of technology and domestication processes.

Hence, with digitization we refer to the on-going development of mobile and virtual communications between hospitals, homes and caring units moving focus from hospital- and function-based organizations to personalisation and new arenas for health care. In other words, from talking about health care in terms of patients, diseases, wards and elimination of risks with single technical applications, towards a focus on health, home health care such as cancer and palliative care, monitoring and communication on distance, digitized and accessible patient care records requiring active patients and citizens and more of inter-professional collaboration and teamwork between health care professionals. These shifts in the way care and contacts with citizens are understood is already underway applied in a wide range of ideas but less supported by empirical evidence.

Digitization will definitely challenge the way engineers and nurses interact, both in real life caring situations but also in developing and implementing digital health care. For digital applications to be supportive in local practice it is dependent on engineering competence but also relevant implementation outside laboratories, in real life settings, and situated caring competence provided by nurses in close collaboration with patients. On the other hand, digitization offers new opportunities to collaborate in areas not yet occupied by any specific interest. The potential to improve collaboration is still embedded in the interaction between technology, user experiences and the way the context is organized.

C. Patient´s new role

One of the most important areas today in which inter-professional and multi-disciplinary collaboration is crucial is the implementation of telemedicine for outpatient care including system design for monitoring, design of care robots and applications for self-care. This development has a number of consequences, primarily a further shift from in-patient care to caring activities outside hospitals and for engineers the challenging transition moving new applications from the laboratory out to practice. The patients themselves have become an indispensable factor in making the systems work, since the implementation of digitized home health care is taken place outside controlled hospital environments and laboratories. This in turn implies understanding the domestication of technology and awareness of the environment and the situation where the systems are supposed to work. In addition, these systems are increasingly complex. The way to understand technological worlds and social worlds has long been perceived as “socio-technical systems. Factors that help coordination and adaption in complex socio-technical systems have been brought forward such as shared visions and common goals [42] and active learning environments [43] [44]. The traditional way to apply already made technology is criticized for being unaware about the social context and the human being exposed to a more or less deterministic development [45] [46]. One reason is that technology evaluated as onedimensional building on a linear model of innovations makes recommendations that are underestimating complexity [41]. Consequently, it is hard to find empirical evidence that implementations benefit from linear systems. On the contrary, it is estimated that two-thirds of organizations' efforts to implement change fails [48]. For these reasons it is suggested that one must leave the linear model and consider all factors that influence implementation [49]. Actually, this is a point where multidisciplinary collaboration is needed based on a more informed and critical perspective on innovation and implementation. From an engineering point of view, mathematicians can contribute to a more non-linear thinking by providing the central distinctions between linearity and non-linearity. A broad range of researchers in caring science and social science can contribute with empirically justified concepts of the role of technology in human contexts by providing concepts such as domestication, behaviors and context.

So far, however, the uptake of digital technologies designed for patient centered care and use implemented in people´s everyday life [50] has remained disappointing. They typically run up against acceptability problems and widespread non-use when they meet the muddled realities of everyday life and complex market forces. Their interventionist potential has not been realized because the complexities are both under-theorized and over-instrumentalized [51]. For children in long-term intensive care the needs for design that allow them to live an active life are crucial. For example, children with respiratory disorders in need of carrying around oxygen equipment at home are still to a great extent discriminated by heavy and stigmatizing design. The same goes for elderly people. The way the most common technologies provided by home health care reflects images that these users have very low expectations. In fact, this can be experienced as stigmatizing [52]. At the same time investments in systems to meet increasing needs in health care to provide help and support for ageing populations aiming at facilitating for elderly people to stay in their own homes as an alternative to nursing homes are huge. European Union invested more than €1 billion euro in research and industry collaborations in long-term monitoring in combination with robots in the homes of elderly people between 2008 and 2013 and continues to do so in Horizon 2020.

D. Digitization goes beyond single applications

Digitization in this paper is framed in the discourse of eHealth, implementation and learning, understanding digitization as leading up to a greater complexity and especially for engineers, a new sometimes puzzling context coined in terms of implementing technology "out in the wild", outside the laboratory. We are entering a new phase where it is more about interconnected systems and no longer just individual applications. Today, the use of different types of IT applications is not unknown to anyone in health care. Examples of products that can add value are digital patient records, alarms, and sensors for monitoring health, robots with various appearances such as social robots and rob cats, digital incontinence indicators, and remote surgery, decision support for diagnosis and balance training for stroke sufferers. Many of these examples have been shown to increase the quality of care and have already become successful business solutions. At the same time, this raises awareness about the fact that individual technical artefacts are hardly the solutions to the health care problems. This leads to a number of questions which need to be addressed: How can we permanently and sustainably integrate new applications in health care? What is the best way to implement accurate solutions in health care with a comprehensive and ongoing digitization?

Another important question concerns what is called the technological imperative in relation to caring values: are we always obliged to do what can be done in terms of technological development? Or can we find ways to criticize such deterministic views? Although technology is closely associated with the development of modern medical care, the relationship to technological development is divided [53]. Here is a criticism that high-powered specialization risks creating problems can become counterproductive. The German philosopher George Henry Gadamer asserted, for example, that it is precisely in highly developed technical civilizations that the phrases "quality of life" and "whole" are expressed, because something is lost. A narrow technical perspective sometimes tends to give healthcare professionals the role of managing technology instead of people. It also contributes to the technological imperative, i.e., what is possible to be measured must be measured even if the benefit is unclear. Meanwhile, with a critical perspective on technology development, we can see its growing importance and that it is a force for change.

To sum up, while engineering understands the increased complexity, patient context is more familiar to nurses. As digitization increasingly moves in to the realms of health and self-care, the relationship between the caregiver and the individual citizen, patient or care receiver, becomes more important. In order to successfully implement and promote self-care management, personalized medicine" and consequences of the demographic development this paper will contribute to open up the "black box" providing new findings from the inside of collaboration and learning activities between engineers and nurses [46] [55].

E. Aim

With this background, how much can we expect engineers and nurses to collaborate and what can be defined as factors leading to successful interactions and what are the obstacles? We can assume that collaborating in order for the patients and citizens to benefit from digitization and be safe and cared for in a relevant way both inside and outside hospitals encompasses both common interests but also potential conflicts. These conflicts can be of a more general nature such as differences in understanding technology and its role in caring situations, power relations in providing and receiving care and how to influence technological developments, as well as more specific context dependent conflicts for example personal relations.

More precisely, how do engineers understand the caring situation and how do nurses understand technology in caring situations and what makes it work? How does this interaction take place in research and development projects and how can this interaction be prepared for and facilitated already during education and training?

This paper aims at deepening the knowledge on collaborations between nursing and engineering; what are the factors leading to successful interactions and collaboration and what are the missing links? What are the challenges considering the practical context in which digitization takes place and asks for successful implementations as a result of multidisciplinary collaborations?

F. The structure of the paper

This first section, which is an introduction to the empirical results presented in the next section, has shown how digitization is characterized more by systems and complexities rather than individual artifacts, and therefore requires enhanced collaboration between engineers and healthcare professionals to get systems to function outside the laboratories and outside hospitals, in the homes of patients and caregivers. Section II provides the empirical result from two cases. The first paragraph presents the result from an in-depth analysis of the collaboration between engineers and nurses in two European based projects on robotics related to health care. The second paragraph presents the evaluation of joint learning activities between engineering and nursing students during their last semester. These activities were taken place as pilots to develop a strategic program in collaboration between Royal Institute of Technology (KTH) and the Swedish Red Cross University College (SRCUC) and the new subject: Technology in Health Care. Both cases are original research elaborating on what are the common interests and existing conflicts between nursing and engineering when implementing and evaluating new technologies in caring situations. Section III discusses this result and methodological weaknesses and strengths related to the cases. Section IV provides conclusions of use for future engineering-nursing collaboration.

II. EMPIRICAL FINDINGS IN TWO CASES

Result from two empirical cases is presented including methodology, research design and result. Targets groups for both cases are engineers and nurses in collaboration and their relation to patients in caring situations.

A. First case: implementing robots and systems in elderly people´s homes in joint collaboration between nurses and engineers

1) Methodology and research design

The first case is a meta-analysis of two research and development projects funded by the European Commission within call FP7-ICT-2011-7. The project GiraffPlus was aiming at the implementation of a system with sensors and a tele robot (half autonomous) for monitoring health care needs and communicating with health care professionals and municipal home help services. The second project, HOBBIT, was aiming at a robot companion to help elderly people at home picking up things from the floor, retrieving glasses or medication or call for help in case of falling. From these projects, which lasted for three years, there are a lot to learn even though the projects in themselves had a limited time to make changes that affected the project plan including project goals and the level of ambition.

The analysis is based on a review of critical situations that emerged during the work and demanded a joint discussion in the project teams or any kind of change of plans. The critical situations that were defined derived from projects combined and their protocols, notes from meetings and from the on-line bug trucker. The critical situations discussed in protocols or in notes from meetings or pointed out as problems in the on-line bug trucker were organized in line with the project plan including developing ethics and ethical vetting; catching needs in scientific literature reviews; translating needs into technological applications; laboratory tests; tests in the homes of elderly people; and exit strategy.

The competencies involved a number of professional actors but for this analysis engineers and nurses and patients were selected as focus points. The projects were organized in different ways. Test sites were organized in seven homes in the HOBBIT project in three European countries and in total 15 homes in three countries in the GiraffPlus project over periods from three weeks up to eighteen months. While HOBBIT were centrally organized with one coordinator for all tasks and engineers moving around to different partner countries with different test sites, the GiraffPlus project were organized in test site teams in each of the three partner countries being responsible for in total fifteen the test sites. The test site teams had regular contact or immediate contact in case of emergencies or technical problem. The test site teams included four functions: managing the project, coordinating the system, contacting users and solving day to day technical problems. Half way into the GiraffPlus project an online bug trucker was installed in the report system managed by the test teams to get a better overview of what kind of problems that the elderly users encountered. The on line bug trucker registered technical bugs and other test site problems i.e. in the homes of the elderly users. In total five engineers and two nurses were part of the HOBBIT project and six engineers and five nurses in the GiraffPlus project.

The elderly participants thus had a central role in the planning of both projects being involved in every step of a user centric development cycle. They were selected as primary users and asked for participation on the basis that they met the following inclusion criteria: being 60 years or older; willing to try out a robot in the home for three weeks in the HOBBIT projects and 6-12 months in the GiraffPlus project; living in their own homes, not in a nursing home or sheltered home for elderly people; being frail with walking instability and risk of falling and feelings of being insecure; at least one chronic condition such as: diabetes, high blood pressure or cardiovascular disease; and being on medical treatment (taking pills). Not having any diagnosis such as dementia or other cognitive disease was an important exclusion criterion. Their participation encompassed tests of a variety of products in a laboratory related to the monitoring system such as blood pressure measurer and remote controls for the robot. Second, they were having a monitoring system with a telerobot in the GiraffPlus project or a mutual care robot in the HOBBIT project installed in their own homes.

2) Result: Common interests and conflicting epistemologies

The results show that critical situations appeared in relation to problem interpretations and to the need to develop and follow ethics approved, catching and translating needs when designing robots and tools, finding test sites, conducting field tests and in the end, withdrawing from the homes, the exit strategy. These critical situations did not always bring about changes but exposed conflicting epistemologies and how they affected implementation and involvement of users. To solve these dilemmas without risking the progress of the project it goes without saying that the participants completed the project being in for different reasons. Four critical situations, or epistemological gaps, were found:

a) Cultivating images of elderly in need of technology

In the first category, we learn something about the validity of the project result, namely that the input engineers needed, despite the user-centered approach were more often guided by their images of elderly rather than of what the older participants in the project expressed. This was confirmed in a master thesis published in August 2017 within this KTH-SRCUC program [56]. The aim of using a user centered design approach was to ensure active involvement of elderly users during the entire development cycle ranging from the analysis, observation, design and verification phase in order to catch the most relevant needs of this target group. This was an important requirement for getting funding for the project. For this reason, their involvement as such was legitimizing the project.

However, to what extent did that lead up to catching relevant needs? One of the first tasks was to map out scientific evidence for elderly people´s needs of home health care systems as a base for making the first questionnaire before testing in the homes. Results of a literature review in the GiraffPlus project was reported in the first delivery to the EC and followed-up twice compared with what the elderly themselves asked for along the way. It turned out that the result from the literature review was not confirmed by the elderly persons who tested the technology in their homes. There was a lack of useful concepts to catch context-based factors and experiences. It turned out that the relation between individual physiological and social needs related to single applications did not match with what the elderly users tried to convey. For this reason, the project ran the risk to continue to be shaped by stereotypic views of elderly rather than broadening the understanding of aging and later life in a technological landscape. Also, the elderly themselves seem to refer to stereotypic views when they most often said that the technological solutions that was presented probably would be useful for other elderly but not for themselves.

b) Being in for different reasons

Second, we learn that engineers and test persons can be in the same project and perform tasks in parallel but with very different reasons. Both the engineers, social scientists and health care professional were motivated to participate in the project out of their scientific perspective or engagement in certain questions. The main objective was to develop or study the communication and information transfer between patients in their homes and their health center (physician, nurse, physio therapist or occupational therapist) using robots and sensors. The goal was to make this contact more efficient, but not necessarily to give more opportunities for personal contacts. However, the reason for the elderly to be a part of the project was exactly that, to gain more contacts with health care and a feeling of being watched over and taken care of. The presence of project people going in and out of their home even became too much for one person who dropped out, but many of them expressed expectations to be more in contact with the health center as a result of the project. To conclude, while the engineers tried to rationalize contacts and ease the burden for health care providers, the elderly users saw this as a way to gain more contacts with health care. It can be added that the ambition of rationalizing contacts would have required re-organizations of doctor's routines and health care organization that was beyond this project.

c) Making robots work

Third, testing communication between the elderly person at home and the robot itself was the main objective for the HOBBIT project. Also for the GiraffPlus project there were possibilities for the elderly person to self-initiate a communication, for example in emergencies, in need for rehabilitation with support from a physio-therapist; or getting virtual visits from a nurse. To create systematic dialogues possible to evaluate, engineers used scripts i.e., predefined dialogues and behaviors, that test persons were supposed to follow. In the GiraffPlus project hardly anyone initiated the robot. In the HOBBIT project tests with script and users focusing a situation were tested in a laboratory, trying out for example the need for entertainment or asking for help picking up something.

What happened in the GiraffPlus was that the wider context interfered; a domestication process took place meaning that the robot became a part of the social context at home including routines, habits and the moral economy of the household [57]. While scripts made them isolated from their social context, the robot became meaningful when it was given a name and became a symbol for the elderly persons being part of the most advanced technological progress. As a consequence, one of the most important aspects for them was to make sure that visitors could see the robot and learn about their participation in the project. From this experience it can be questioned if the script tested in the lab would be possible to be implemented in the home in the way it was expected. To conclude, while the engineers were trying to predefine dialogues and behaviors, the domestication process that took place in the context of private homes alternated the result given in a laboratory. With domestication, elderly users came to use and evaluate the technology in different ways. An assessment of how useful the new technology was in relation to what they already had led to lower or higher usage. If it was easier to call the phone instead of getting connected to the telerobot this is what they did. Another example is how visible one wanted the new technology to be at home, which had consequences for where it was placed. Another example was how well the new technology could be adapted to daily routines. One refused to use it if it was inconvenient and made everyday life more complicated or even intrusive. This might be the most obvious proof of the difference between controlled laboratories and people's social world and the factors which affects implementation.

d) Making them ours

Fourth, the project aims to define relevant applications and make technology work was overshadowed by the fact that the elderly participants put great emphasis on the robots' appearance. The initial workshops about what to use a robot for and its design and appearance engaged groups of elderly persons invited to draw pictures and discuss these matters. Already at these workshops it was obvious that the appearance of the robots engaged more compared with technological functions. As part of the domestication process these artifacts turned into family members. Other signs of domestication were that the robots were given names and that the elderly users tried to fit the system as much as possible into their daily life, meaning as few changes as possible. One example was the refusal to be micro-managed, i.e., that their behaviors were monitored just for the interest of the project.

Three design briefs of potential robot appearances were developed and tested in three countries. With reference to the project's limited time and funding it was not possible to test a variety of prototype appearances, but the picture was commented on by the elderly as an important part of the interior style of their homes. One can assume that this affected their use of the robot and their assumptions about having such a "machine" at home. With this background, one plausible assumption is that while the engineers were focusing technology, the elderly users focused on the looks of the robots.

B. Second case: joint learning activities for engineers and nurses preparing for technological hazards

1) Methodology and research design

The second case is about preparing biomedical engineering students and nursing students for collaboration in caring situations in which these two professional groups will meet and handle any risk and hazard that might occur when caring for patients. Testing this joint learning initiative has two aims: to improve communication between them and to make them more proactive in, not only avoiding risks but also in taking part of technological development and collaborate around producing new applications and innovations to secure patient safety.

This learning activity took place within the collaboration between The Royal Institute of Technology, KTH, and Swedish Red Cross University College, SRCUC, in Stockholm Sweden, implemented in parallel with a new interdisciplinary postgraduate program and teaching subject - Technology in Health Care (in Swedish: Teknisk vårdvetenskap). The Nursing Science program at SRCUC will therefore have the potential to add a unique technological profile to its 150-year tradition of training nurses. At KTH, the Department for Technology and health will be given new opportunities to implement new technologies in health care successfully and to improve its research on how technology works in caring settings. Three goals was set in 2014, in the beginning of the strategic development: 1) to better understand medical technology, safety aspects and functions; 2) to increase the ability to proactively participate in the development, implementation and evaluation of technology; 3) to understand how technical developments affect professional roles and working methods. One strategic meeting point early defined was between engineering and nursing students.

Preparations for learning activities of which one is presented in this paper include publications focusing patient safety, learning and innovation. The first articles in this context were published by Mattsson & Stevens [54] [58] and Björling on coated endotracheal tubes and central venous catheters with focus on patient safety [60] [61]. During the same period Östlund published several articles on innovation processes, design and ethics especially dedicated to the use of robotics in elderly care [62] [63] [64].

The first attempts to joint learning opportunities for engineers and nurses took place in 2016 and was followed up 2017, conducted in collaboration between teachers from KTH and SRCUC. The exercise is organized as a part of the curriculum in which students in nursing and students in medical engineering meet for a joint activity during their last and sixth semester of education, in total 80 students participated. The starting point for the joint exercise was the caring situation where they both have their professional roles but rarely access to each other's perspectives. The education goals were to increase the understanding of each other's way of approaching a caring situation with advanced medical technology, more precisely to increase the understanding of each other's perspectives, competence, tasks and way of communication; to start to communicate with each other; and to solve a problem that is supposed to be as realistic as something they will approach as part of their future work situation.

The work material consisted of a report from the Swedish Accident Investigation Authority about a serious failure at a Swedish hospital causing the death in the cardiac intensive department in 2010. This accident was the object for an investigation published in an official report [65]. The students prepared themselves before the exercise by reading this report. They were also prepared with the theoretical model of a "Man-Machine-Organization" model which were a part of both nursing and engineering educations. The exercise took place during one day and gave them the opportunities to elaborate on the education goals in groups, presenting the result of the group work for the entire group and to make an individual statement of the learning outcomes.

A total of 59 students participated in the workshop in 2016, whereof 39 engineering students and 20 nursing students. When giving the same workshop with a new set of students in 2017 a total of 69 students participated, whereof 39 engineering students and 33 nursing students. At both workshops all students were asked to voluntarily hand in an anonymous written reflection about the joint workshop described above. They were asked to evaluate the activity by responding three questions: 1. What are your views on patient safety after today's workshop? 2. What are your views on collaboration between biomedical engineers and nurses after today's workshop? 3. What did you learn from today's workshop? All students handed in their evaluations.

2) Result

The result of this joint learning activity gave overall very positive result, which is a conclusion drawn from anonymous written evaluations the same day as the exercise took place. The most referred experiences were that this was educational and surprisingly interesting to deepening the understanding of nursing respectively engineering perspectives. Common insights were that technology is an important support but dependent on the way it is organized and that there are no unrelated actions that does not have consequences in such a situation. Critical comments asked for more preparations and the need for follow ups out there in real working life.

If we consider how this activity affects long-term collaboration between engineers and nurses, we can say that the result proves that they have gained new insights in professional jargons and not least the awareness that there are other professions involved in making the caring situation into what it is. Both these groups will in a few months meet out in the labour market. They have no problems to find employments since there is a shortage of both nurses providing care and engineers working at hospitals with providing equipment and support for surgery and a wide range of follow-up health care activities within the responsibility of hospitals. But at that time it might not be possible to create the communication that in this case is provided already during the education. The possibility for joint learning activities between nursing students and medical engineers during their education will broaden their understanding, enhance patient safety and ensure a sustainable care.

III. DISCUSSION

This paper is looking for factors leading to successful interactions and collaboration between engineers and nurses and what are the missing links. While such collaboration is taken for granted as something that will automatically lead to more useful technologies and implementations and no one opposes such an approach it is rare to find explanations to why it works or does not work. The results reported in this paper point to the lack of a common understanding of what creates successful implementation and why the results of testing of products in laboratories cannot automatically be transferred to real life settings without taking into account the user´s expectations, skills and social contexts. This is described in the analysis as engineers and nurses have different epistemologies.

Engineers expect laboratory testing to be a reliable method to make technological products work outside laboratories, while nurses are expected to learn to use the technological devices that they are provided with by engineers. Both groups under-estimates the domestication processes that takes place as a result of the interaction between new technologies and user´s in real life contexts and interactions happening in caring situations. Nurse´s experiences of using technologies in caring situations was not counted as part of their technological competencies or at least not asked for when prototypes were developed and testing was planned. Neither did the nurses see themselves as part of the technological development. It was the engineer´s responsibility. The nurses relied on published result on what factors are of importance for independent living in relation to daily activities.

The key user group – the elderly users involved in the projects described in this paper – constitutes a third collaborator that also brought in their expectations and which turned out to be quite different from the nurse´s and engineer's expectations. The elderly users can be described as being "implicated" users or "lay end user", meaning that they are talked about and involved in responding to interview questions and even observed in the home, but still not involved in modifications or design [66]. It seems as if engineers, nurses and elderly users can be involved in the same development project but for very different reasons. It can be assumed that there are hidden aspects not properly explored in this design process that might be innovative. The relationship between these three collaborators could for example be studied in terms of power and who has the advantage of formulating problems and solutions. That would probably reveal more of these mechanisms of why they were in for different reasons.

Besides the four critical situations that are described in this paper and that reflect the phenomena of being in for different reasons there were also other findings pointing at elderly user's context being invisible even though they were supposed to be at the heart of the project. They were aspects related to safety and security. Since security is well linked to transfer of information and private data, more or less unknown aspects occurred with the implementation of technologies in the homes, such as the capacity of the infrastructure that caused severe interruptions; the interios design and furnishing not adapted to robots moving around; the social context having its own routines which was interrupted by the system and sometime experienced as intrusive.

Digitizing health care places new demands on a long-established organization of engineers being employed at the medical technology unit and nurses who have their main workplace with patients at the ward. With digitization it is no longer enough for engineers to provide and manage the technology at the hospital and nurses to learn to use individual machines. Now health care is taken place with new involvements of patients and at new arenas in people´s homes and on the move. The result challenges the way innovation processes are normally organized as linear, putting two and two together under limit time pressure and predictable outcomes. For engineers this will probably widen their scope from the hospital to environments outside the hospital and for nurses it will generate new tasks and new professional roles such as partake in distance surgery, keeping in touch with patients at home via robots or being genetic guides for patients finding out about their health heritage. Already comprehensive investments are in motion where engineers and nurses are supposed to collaborate around design and implementation and tests.

Being in for different reasons tells us that a deeper analysis of the actual collaboration and critical situations in joint projects show where the limitations are. It is not about testing the design of interfaces only or pre-decided effects of using certain devices. Neither is the result of implementing lab tested technology in real life settings predictable. What makes a difference are factors such as the expectations of user's ability, very often concluded from generalized user requirements leaving out the context. Second, user's own expectations, in this case they were not interested in the robot per se but to be involved in frontier project and increase contacts with health care. Third, it is what´s in it for me that makes the domestication process successful, not technological imperative such as monitoring behavior and involving relatives just because it is possible. These findings challenges both engineering and nursing paradigms since it is not about patients only or not about technology only but about the interaction and the caring situation.

These findings should be of interest for policymakers and planners of research and development programs on national and European levels when elaborating on what should be the criteria for investments in future home health care systems. In connection with the discussion of an aging population and the potential of technology, these results are important in understanding what can make a difference. For engineers these results can help to take their technologies a step further by closing the gap between the laboratory and real life settings, sometimes called "out there in the wild" as some robotic engineers sometimes call this world where people interact with their result.

Some of these findings can be generalized to health care professionals beyond nurses while other findings are specifically relevant for nurses and for the caring situation. Being close to the patient, following in detail the patient's caring needs provides a processual understanding of the development and a deeper sense of tacit knowledge compared to temporary consultations or surgeries.

How to teach technology in nursing is a comprehensive question, not least educational. This discussion is taken place within the teachers group, appointed to be responsible for teaching aspects of technology. The second case on joint learning activities is based on two pedagogic ideas: not starting with change but first understanding what is already in motion (engineers and nurses side by side in a caring situation), and second, present ideas that are a truer expression of what kind of knowledge these two groups need to be prepared for. Since technology is always in progress finding ways to collaborate and communicate independent of what kind of technological changes be considered to be more sustainable than inventing the wheel for every new innovation.

Even though the joint learning activities show promising result, to plan for this becoming a permanent part of nursing and engineering education programs is not without difficulties. These education programs are organized based on deeply rooted perceptions of what nurses and engineers need to learn. Changes in curriculum require clear initiatives and can conflict with both space and focus.

Beside learning activities in bachelor education the KTH-SRCUC program also include PhD training and research cooperation, not least joint seminars. One suggestion leading the attempts to find joint research questions is to meet within non-invested areas. A non-invested area is defined as a context or an intellectual discourse where none of the collaborators has invested interests. At least it can be expected to decrease the level of conflicts that can occur in between paradigms. One way of creating a non-invested area is when groups meet that have not met before as in the second case presented above. Another option is to create spaces for collaboration with care providers outside the education or research department. To fill these spaces with a content, depend on the needs and demands of the care providers and requires a thorough completion of initial dialogue currently underway.

These two cases presented in this paper are selected out of several attempts to create the KTH-SRCUC program. The strengths of the cases are their unique approach of deepening the understanding of what makes collaboration possible or not possible. It opens windows for new ideas and innovations. The weakness of the first case is that the analysis is made after the completion of the project, not planned on beforehand, which could have included a more systematic collection of different kind of information data. In the second case evaluations were conducted but not yet published in detail. In this paper we draw conclusions in accordance to the aim at deepening the knowledge on collaborations between nursing and engineering by taking on a new perspective pointing out the factors leading to successful interactions and collaboration and what are the missing links.

IV. Conclusions

Collaboration between nursing and engineering has long been requested. This initiative suggests that it takes more than just adding two and two together. There is a need for renewed views on what drives technical development in care and how it can be adapted in a socio-technical system. From previous experiences we learn that preparations for collaboration must include the awareness of epistemological differences as well as common interests to critically examine the understanding of how caring practices are constructed and implemented. For nurses to be proactive requires knowledge about technological developments and the ability to collaborate with engineers and participate in design and innovation processes both for healthcare professionals and concerned citizens. For engineers a more thorough understanding of caring situations and users will contribute to a more reliable provision of developed solutions and point at new ideas leading up to innovations.

References

[1] Östlund, B. Björling, G. Mattsson, J., Stridh, S. & Sahlström. M. Technology in Health Care. A new research and teaching subject in collaboration between nursing science and engineering. Ninth International Conference on eHealth, Telemedicine and Social Medicine, eTelemed 2017 March 19-23.

[2] J. Frenk, L. Chen, Z. A. Bhutta, J. Cihen, N. Crisp, T. Evans, H. Fineberg, P. Garcia, Y. Ke, P. Kelley, B. Kistnasam, D. Naylor, A. Pablos-Mendez, S. Reddy, S. Scrimshaw, J. Sepulveda, D. Serwadda and H. Zurayk. “Health professionals for a new century: transforming education to strengthen health systems in an interdependent world,” *The Lancet* 376, No. 9756, 2010: 1923-1958.

[3] I. R. Hallberg, “Moving nursing research forward towards a stronger impact on health care practice,” *International Journal of Nursing Studies*, 46, 2009: 407-412.

[4] *Swedish Society of Medicine*, Onödiga eller skadliga åtgärder i svensk allmänmedicin – ett diskussionsunderlag, [Unnecessary or harmful actions in Swedish general medicine - a basis for discussion], Stockholm: Swedish Society of Medicine, 2013.

[5] *Swedish Society for Nursing*:

- Kunskapsluckor inom omvårdnad. En första inventering [Gaps within caring. A first inventory], Stockholm 2013.

- Strategi för sjuksköterskors arbete med eHälsa [Strategy for nursing and eHealth], Stockholm 2012.

- Värdegrund för omvårdnad [Valuebased caring], Stockholm 2010.

[6] S. Bakken, “Informatics for patient safety: a nursing research perspective,” *Annual Review of Nursing Research*, 24, 2006: 219

254.

[7] M. E. Porter and E. Olmsted Teisberg, “Redefining Health Care. Creating Value-Based Competition on Results,” *Harvard Business Review Press*, 2006: 98-148, 2006.

[8] L. Kurzweil, “2045: The Year when Man Becomes Immortal”. *Time Magazine* February 10, 2011.

[9] C.P. Snow, “*The Two Cultures. And a second look”.* London: Cambridge University Press, 1964.

[10] D.J. Mick, and M.H. Ackerman, “Nursing and biomedical engineering transdisciplinary clinical trials collaboration”. *Medical Devices*, 2, No. 2, 2005: 131-133.

[11] C. Hurlock-Chorostecki, C. Forchuk, C. Orchard, S. Reeves, and M. van Soeren, “The value of the hospital-based nurse practitioner role: development of a team perspective framework”, *Journal of Interprofessional Care*, 27, No. 6, 2013: 501-508.

[12] K. Perkin, “Nurse practitioners and interprofessional collaboration”, *Journal of Interprofessional Care*, 25, No. 4, 2011: 243-244.

[13] J.D. Gonzalo, B.M. Thompson, P. Haidet, K. Mann and D.R. Wolpaw, “A Constructive Reframing of Student Roles and Systems Learning in Medical Education Using a Communities of Practice Lens”, *Academic Medicine*, 20, No. 10, 2017

[14] B.F. Brandt, F.B. Cerra, and C.W. Delaney, “The United states Center for Interprofessional Practice and Education: Integrating an informatics approach to interprofessional work”, *Journal of Interprofessional Education*, 29, No. 6, 2015: 592-595.

[15] P. Hall, “Interprofessional teamwork: professional cultures as barriers”, Journal of Interprofessional Care, 2005: 188-196.

[16] F. Abdelrazek, B. Skytt, M. Aly, M.A. El-Sabour, N. Ibrahim, and M. Engström, “Leadership and management skills of first-line managers of elderly care and their work environment,” *Journal of Nursing and Management*,18, No.6, 2010: 736-745.

[17] H. Hagerman, M. Engström, E. Häggström, B. Wadensten, and B. Skytt, “Male first-line manager’s experiences of the work situation in elderly care: An empowerment perspective,” *Journal of Nursing Management*, 23, No. 6, 2015: 695-704.

[18] L.W. Samia, C.H. Ellenbecker, D.H. Friedman, and K. Dick, “Home Care Nurses Experience of Job Stress and Considerations for the Work Environment,” *Home Health Care Quarterly*, 31, no.3, 2012: 243-265.

[19] T. Suominen, N. Savikko, K. Kiviniemi, D. I. Doran, and H. Leino-Kilpi, “Work Empowerment as Experienced by Nurses in Elderly Care,” *Journal of Professional Nursing*, 24, No. 1, 2008: 42-45, 2008.

[20] K.G. Davis and S.E. Kotowski *“Prevalence of Musculoskeletal Disorders for Nurses in Hospitals, Long-term Care Facilities, and Home Health Care: A*

Comprehensive Review," Human Factors, 57, No. 5, 2015: 754-792.

[21] M. Simon, P. Tackenberg, A. Nienhaus, M. Estryn-Behar, and C.P. Maurice, H.M., Hasselhorn, "Back or neck-pain-related disability of nursing staff in hospitals, nursing homes and home care in seven countries-results from the European NEXT-study," *International Journal of Nursing Studies*, 45, No. 1, 2008: 24-34.

[22] V. T. Elst, C. Cavents, K. Daneels, K. Johannik, E. Baillien, A. Van Den Broeck, and L. Godderis," Job demands-resources predicting burnout and work engagement among Belgian home health care nurses: A cross-sectional study," *Nursing Outlook*, 64, No. 6, 2016: 542-556.

[23] R.R.M. Gershon, M. Pogorzelska, K.A. Qureshi, M. and Sherman, Home health care registered nurses and the risk of percutaneous injuries. A pilot study," *American Journal of Infection Control*, 36, No.3, 2008: 165-172.

[24] A. Hogh, H. Giver, H. Hannerz, B. Pedersen, and B. Holbæk "Bullying or violence during training snd the risk of dropout from the eldercare sector 2 years later," *Scandinavian Journal of Caring Sciences*, 26. No.4. 2012: 730-737.

[25] J.K. Leiss, J.T. Lyden, R. Mathews, K.L. Sitzman, A. Vanderpuije, D. Mav, M. Kendra, M.A. C. Klein, and C. J. Humphrey, Blood exposure incidents rates from the North Carolina study of home care and hospice nurses," *American Journal of Industrial Medicine*, 52, No. 2, 2009: 99-104.

[26] P. Markkanen, C. Galligan, A. Laramie, J. Fisher, S. Sama, and M. Quinn, "Understanding sharp injuries in home healthcare: The Safe Home Care qualitative methods study to identify pathways for injury prevention. "*Environmental and Occupational Health"*, 15, No.1, 2015.

[27] R.Rugulies, I.E.H. Madsen, P.U. Hjarsbech, A. Hogh, V. Borg, I.G. Carneiro, and B. Aust, "Bullying at work and onset of a major depressive episode among Danish female eldercare workers," *Scandinavian Journal of Work, Environment and Health*, 38, No.3, 2012: 218-227.

[28] J. Gené Badia, S. A. Borràs, C.S.J. Carles, C. Camprubí, M Dolors ; F. Cegri Lombardo, T. Heras Antonio ; R.R. Noguera, S. Martinez González, A. Oliver Olius, M. Piñeiro González, T. Montanuy Baró, Teresa, E. Limón Ramírez, M. A. Aranzana L. R. Pedret, M. M. Borrell, G.A.Hidalgo, "Nursing workload predictors in Catalonia (Spain): A home care cohort study," *Gaceta Sanitaria*, 25, No. 4, 2011: 308-313.

[29] A.Larsson, L. Karlqvist, M. Westerberg, and G. Gard, "Identifying work ability promoting factors for home care aides and assistant nurses," *BMC Musculoskeletal Disorders*, 13, 2012.

[30] G.B. Morgan," Job satisfaction in the home health care context: Validating a customized instrument for application," *Journal of Healthcare Management*, 55, No.1, 2010: 11-21.

[31] C. Westermann, A. Kozak, M. Harling, and A. Nienhaus, "Burnout intervention studies for inpatient elderly care nursing staff: Systematic literature review," *International Journal of Nursing Studies*, 51, No.1, 2014:63-71.

[32] A. Nilsson and M. Engström "E-assessment and e-training program among elderly care staff lacking formal competence: results of a mixed-methods intervention study," *BMC Health Services Research*, 15, 2015.

[33] C.D.N. Rasmussen, A. Holtermann, O.S. Mortensen, K. Sagaard, and M.B. Jargensen," Prevention of low back pain and its consequences among nurses 'aides in elderly care: a stepped-wedge multi-faceted cluster-randomized controlled trial," *BMC Public*, 13, 2013.

[34] C.D.N. Rasmussen, N.K. Lindberg, M.H. Ravn, M.B. Jørgensen, K. Søgaard, and A. Holtermann," Processes, barriers and facilitators to implementation of a participatory ergonomics program among eldercare workers," *Applied Ergonomics*, 58, 2017: 491-499.

[35] G. Syson and J.Bond "Integrating health and social care teams in Salford," *Journal of Integrated Care*, 18, No. 2, 2010: 17-24.

[36] I. Vedel, M. De Stampa, H. Bergman, J. Ankri, B. Cassou, F. Blanchard, and L. Lapointe, "Healthcare professional's ad managers´ participation in developing intervention: A pre-intervention study in the elderly care context," *Implementation Science*, 4, 2009.

[37] V. Sorlie, L. Jansson and A. Norberg, "The meaning of being in ethically difficult care situations in paediatric care as narrated by female Registered Nurses," *Nordic College of Caring Sciences*, 17, 2003: 285-292.

[38] L. Fagerstrom, K. Eriksson and I. B. Engberg, "The patient's perceived caring needs: measuring the unmeasurable," *International Journal of Nursing Practice*, 5, No. 4, 1999: 199-208.

[39] P. E. Benner, C. A. Tanner and C. A. Chesla, *Expertise in nursing practice : caring, clinical judgment & ethics*, 2d. ed. New York: Springer Pub. 2009.

[40] J. Mattsson, M. Forsner, M. Castrén and M. Arman, "Caring for children in pediatric intensive care units: An observation study focusing on nurses' concerns," *Journal of Nursing Ethics*, 20, No. 5, 2013: 528-538.

[41] E.M. Rogers, "*Diffusion of innovations"* 4th.ed., New York: Free Press, 1995.

[42] J.C. Gorman, N.J. Cooke, and E. Salas,"Preface to the Special Issue on Collaboration,Coordination, and Adaptation in Complex Sociotechnical Settings," *Human Factors Journal Special Issue on collaboration, coordination, and adaptation in complex Sociotechnical Systems*, 52, No. 2, 2010: 143-146.

[43] J. Y. Mattsson and L. Stevens, "Development of an individual assessment instrument for critical care nursing students," *Journal of Nursing Education and Practice*, 7, No. 2, 2017: 54-61.

[44] H. Engstrand and J. Mattsson, "The non-verbal communication in handover situations are the spice between the lines, to understand the severity of the patient´s condition," Journal *of Nursing Education and Practice*, 7, No. 5, 2017: 1-8.

[45] T. Greenhalgh, F. Macfarlane, P. Bate, and O. Kyriakidou. Diffussion of innovations in service organizations: systematic review and recommendations. *Milbank Quarterly*, 82: 2004: 581-629.

[46] Damschroeder, D.C. Aron, R.E. Keith, S.R. Kirsh, J.A. Alexander & J. Lowery. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, 4, No. 50, 2009.

[47] E. Callahan Hunt, S. Breckenridge Sproat, and R. Rutherford Kitzmiller, (2010) "*Nursing Informatics Implementation Guide".* New York: Springer, Health Informatics Series.

[48] S. Koch, M. Hägglund & I. Scandurra. *Informatics and Sociotechnical Challenges when Designing Solutions for Integrated eCare*. [Retrieved from www.igi-global.com. 2018-02-13]

[49] L. Huniche & F. Olesen (eds), "*Teknologi i sundehetspraksis"*. Kobenhavn: Munksgaard, 2014.

[50] K. Brittain, L. Corner, L. Robinson, L, and J. Bond, "Ageing in place and technologies of place: the lived experience of people with dementia in changing social, physical and technological environments". *Sociology of Health & Illness*, 32(2), 2010:272-287.

[51] A. Peine, A. Faulkner, B. Jæger, and E. Moors, "Science, technology and the 'grand challenge' of ageing—Understanding the socio-material constitution of later life. *Technological Forecasting and Social Change*, *93*, 2015: 1-9.

[52] S. Frennert, "*Older People and the Adoption of Innovations. A study of the expectations on the use of social assistive robots and telehealthcare systems"*. Lund University: Faculty of Engineering, Division of Rehabilitation Engineering, Licentiate thesis, 2013.

[53] H. Eriksson, "*Vårdvetenskap för en postmodern tid*" [Caring Sciences for a Postmodern Era], Lund Sweden: Studentlitteratur, 2014: 29-48.

[54] C. Frostell, G. Björling, E. Strömberg, S. Karlsson, and R.E. Aune. "Tracheal implants revisited". *Lancet,* 389;2017: 10075- 1191.

[55] I. R. Hallberg, "Omvårdnadsvetenskapen i ett framtidsperspektiv [Caring science in future perspectives]," In E. Hamrin, M. Kihlgren, A. Rinell Hermansson and G. Östlinder (eds.), "*När omvårdnad blev vetenskap".* De första decennierna [When caring became science. The first decades], Stockholm: Liber AB, 2014: 328-337.

[56] B. Fischer, "Configuring Users in Laboratories: Elderly Users and Robots". *Master Thesis GEO-2239X*. Utrecht University and KTH, 2017.

[57] S. Frennert, H. Eftring, and B. Östlund, "Case Report: Implications of Doing Research on Socially Assistive Robots in Real Homes," *International Journal of Social Robotics,* 9, No. 3, 2017: 401-415.

[58] J. Y. Mattsson and L. Stevens, "Development of an individual assessment instrument for critical care nursing students," *Journal of Nursing Education and Practice*, 7, No. 2, 2017: 54-61.

[59] H. Engstrand and J. Mattsson, "The non-verbal communication in handover situations are the spice between the lines, to understand the severity of the patient´s condition," Journal *of Nursing Education and Practice*, 7, No. 5, 2017: 1-8.

[60] G. Björling, D. Johansson, L. Bergström, S. Jalal, I. Kohn, C. Frostell and S. Kalman, S, "Tolerability and performance of BIP endotracheal tubes with noble metal alloy coating – a randomized clinical evaluation study," *BMC Anesthesiology*, 15, 2015.

[61] G. Björling, D. Johansson, L. Bergström, J. Sanchez, A. Strekalovsky, C. Frostell and S. Kalman, "Central Venous Catheters Coated with a Noble Metal Alloy – A Randomized Clinical Pilot Study," Acta Anaesthesiologica Scandinavica, 61, No. 8, 2017: 1049-1049.

[62] S. Frennert and B. Östlund, "What happens when seniors participate in new eHealth schemes?" *Disability and Rehabilitation: Assistive Technology*. July, 2015: 1-9.

[63] B. Östlund, Säker vård i hemmet [Patient safety at home] In: M. Ekstedt, and M. Flink, *Hemmet som vårdplats – säker hemsjukvård* [The home as a safe place for home health care]. Stockholm: Liber Publishing, in print, 2018.

[64] S. Frennert, H. Eftring and B. Östlund, "Case Report: Implications of Doing Research on Socially Assistive Robots in Real Homes," *International Journal of Social Robotics*, 9, No. 3, 2017: 401-415.

[65] Swedish Accident Investigation Authority, "Slutrapport Dödsfall på hjärtklinikens hjärtintensivavdelning N25 vid Karolinska Universitetssjukhuset i Solna, Stockholms län, den 17 oktober 2010" [Final report on cardiac failure at Karoloniska Hospital Solna, Stockholm, October 2010], Stockholm: *Swedish Accident Investigation Authority, Report 02, 2013.*

[66] N. Oudshoorn and T. Pinch, 2003. "Introduction: How Users and Non-Users Matter." In *How Users Matter: The Co-Construction of Users and Technology*, edited by N. Oudshoorn and T. Pinch, 1–25. Cambridge, MA: MIT Press.

Improving the Codification of Hospital Discharges with an ICD-9-CM Single-page Application and its Transition to ICD-10-CM/PCS

Cecília Coimbra, Marisa Esteves, Filipe Miranda, Filipe Portela, Manuel Filipe Santos, José Machado * and António Abelha
Algoritmi Research Centre
Department of Informatics, University of Minho
Braga, Portugal
e-mail: {id6881, id6884, id6883}@alunos.uminho.pt, {cfp, mfs}@dsi.uminho.pt, {jmac, abelha}@di.uminho.pt
* Corresponding author

***Abstract*—In recent years, in *Centro Hospitalar do Porto* (CHP), there has been felt an increasing need for a computerized clinical coding tool to aid in the codification of the episodes of hospital discharges from patients admitted to its healthcare units. The process was slow and performed manually by the coding professionals, so there was neither the centralization nor the unification of the information and processes associated with the clinical coding of a hospital discharge. Hereupon, in the context of this study, the aim of the present work was to design and develop a clinical coding tool for ICD-9-CM to support the clinical practice in healthcare units. It additionally included its subsequent transition to the newer ICD-10-CM/PCS coding version. In short, the codification of hospital discharge processes enables the grouping of episodes into diagnosis-related groups (DRGs). The main motivation for the implementation of this classification system is that it provides a financial and patient classification system to contain the costs and waste associated with healthcare services. Thereby, a single-page application (SPA) for ICD-9-CM was designed in order to help health professionals of CHP in their daily work, namely the clinical coding of the episodes of hospital discharges, and it was subsequently updated to the ICD-10-CM/PCS coding version that predominantly improved specificity in describing clinical situations. The main advantages and contributions of the development and use of this Web application are the centralization of information and tasks associated with the coding of hospital discharges, the increase of productivity, and the reduction of wastes of time. Consequently, the ambition is sought to mainly improve the quantity and the quality of work performed by coding professionals at CHP.**

Keywords—ICD-9-CM; ICD-10-CM/PCS; Hospital Discharge; Diagnosis-related Group; Single-page Application; Proof of Concept.

I. INTRODUCTION

The health sector represents a delicate situation for the professionals and the systems responsible for the storage and the processing of clinical data. The main problem with those processes is not in the lack of data but in the diversity and the complexity of the healthcare sector. Since an hospital offers a wide range of services for each patient and clinical condition, it leads to the creation of hundreds or even thousands of specific and unique situations for each one of them. It is then of utmost importance to find a way to measure a hospital productivity and aggregate the multiple activities performed at a healthcare unit.

This specific situation led to the adaptation of the DRGs to this scenario. These kind of classification systems rely on a prior coding system that translates all the diagnoses, procedures, external causes, and morphologies into universal codes like ICD-9-CM and ICD-10-CM/PCS, increasing the semantic interoperability and highly reducing the ambiguity of a discharge report. DRGs are clinically coherent and similar groups that are expected to use the same level of hospital resources.

The introduction of new Web frameworks and solutions for Web development resulted in a new wave of codification platforms for these types of coding systems. With modern layout and intelligent helping tools, it is possible to reduce considerably the codification errors, but also increasing the efficiency associated with the realization of those processes.

Therefore, this paper presents an insight into the development, implementation, and impact analysis of a Web application directed to the ICD-9-CM codification in a major Portuguese hospital located in the north of the country – CHP [1]. On the other hand, it also includes its subsequent update to the newer ICD-10-CM/PCS coding version. This scientific research has been undergoing since the end of 2016.

Thereby, the focus of this paper is mainly to highlight the differences between the prior method (by hand) and the new one with a computerized clinical coding tool, and the transition of the system from the ICD-9-CM coding version to ICD-10-CM/PCS.

In Section II, the state of the art and similar works related to this topic are described. Thereafter, in Section III, the research methodologies adopted are presented in detail. Section IV – “Single-page Application” – presents the work developed and its main results regarding the two versions of the Web tool, followed by a brief Strengths Weaknesses Opportunities and Threats (SWOT) analysis in Section V. In Section VI, the conclusion and future work conclude briefly this paper.

II. STATE OF THE ART

The present section intends to highlight the main theoretical topics addressed throughout this manuscript, and the theory behind the realization of this work, as well as the main studies from the scientific community regarding medical codification, including the subsections “Diagnosis-related Groups”, “ICD-

9-CM Clinical Coding", "ICD-10-CM/PCS Clinical Coding", and "Interoperability: AIDA and AIDA-PCE".

A. Diagnosis-related Groups

A patient classification system is a method in which the main objective is to group patients or disease episodes in order to make it possible to identify their similarities and differences, and therefore allowing that those who belong to the same class are treated similarly.

In this context, DRGs consist of a patient classification system of patients hospitalized in acute hospitals that was developed in response to the rising costs and waste in the healthcare industry [2]–[4]. It groups patients into classes that are clinically consistent and similar in terms of resource consumption [2], [5]. Developed at the Yale University in the United States of America (USA) in the 60's, it is used since 1983 by Medicare to calculate the compensation in cases of hospitalization [6].

This classification system allows defining the set of goods and services that each patient receives according to his needs and the pathology that led to his hospitalization, as well as the defined treatment process. Thus, it is possible to relate the type of patients treated with the resource consumption [2]–[4], [7].

On the other hand, the concept of clinical coherence defines that the pathologies of the patients included in each DRG are related to an organ or system, or even with the etiology, and that the care provided is similarly the same for all the patients in that DRG [7]. A predetermined amount of money is disbursed to hospitals for the treatment of patients belonging to a given DRG, regardless of actual costs associated with the healthcare services provided to them [3], [4], [6], [8].

Thereby, the main motivation behind the grouping of patients from health institutions into DRGs is that it provides a financial and patient classification system that uses the diagnoses, surgical interventions, age, gender, destination after discharge, and other related factors, as grouping criteria [2], [4], [8]. They were introduced in several countries, including Portugal, as a strategy for cost containment, planning, budgeting, management, and follow-up of the healthcare services provided to patients, reducing the disparities and errors [3], [6], [7].

The DRG requires a minimal dataset (MDS) in order to attribute one of the 25 main diagnosis categories to the discharge report [2]. So, the MDS includes, as follows [9]:

- The main diagnosis responsible for the patient admission;
- Other diagnoses;
- Procedures performed on the patient during the internment;
- Gender, age, and height;
- Destination after discharge (transferred, death or discharged against medical order).

Each DRG group has an associated relative weight and weighting coefficient, as well as an exception threshold for the number of hospitalization days that helps convert each case into equivalent patients [9].

Wilm Quentin and colleagues in "Hospital Payment Based on Diagnosis-related Groups Differs in Europe and Holds Lessons for the United States" highlights the differences between the original DRG and the one that countries like France, England or even Portugal implement. This adaption is the basis of most European countries method to finance hospitals, proving to be less cost worthy with a high quality of services [3].

Carina Fourie et al. present in "Systematically Evaluating the Impact of Diagnosis-related Groups on Healthcare Delivery: A Matrix of Ethical Implications" a study of ethical implications and importance of the DRGs in diverse Swiss hospitals [6].

On other hand, in order to highlight the diversity of the subject, Yantao Xin presented a comparison of the amount of medical waste generated in major healthcare units using as basis the DRGs [4].

B. ICD-9-CM Clinical Coding

For the purpose of coding hospital discharges in terms of diagnoses and procedures, in order to allow the subsequent grouping of those episodes in DRGs, the clinical coding "ICD-9-CM" is applied [8]. In short, it consists of a set of diagnosis and procedure codes used for the classification and coding of the morbidity and mortality information for statistical purposes, and for the indexing of hospital records by disease and surgical interventions. The information is then used for storage and research purposes.

Thus, the DRG classification process is performed by a coding professional who must know the system and the classification structure of the ICD-9-CM clinical coding, understand the organization of the indexes and their use in the coding of diseases and procedures, as well as how to apply correctly the principles and rules of the clinical coding ICD-9-CM [8].

Each time a patient is discharged from a healthcare unit, a discharge report is issued from the daily logs by the physician in charge of the patient. The ICD-9-CM clinical coding is a perfect fit to encode diagnoses, medical procedures, external causes, and morphologies, consisting in a universal list of codes recognized in any country across the world.

As the name states, it is an adaptation of the ICD-9 codification system defined and implemented by the U.S. Department of Health in collaboration with the Medicare and Medicaid Service Centres [10]. With more than 13,000 diagnoses and 3,500 procedures, it is essential to develop tools or systems focused on the codification process.

Some Web applications like "Find-A-Code" or the work of Marisa Teresa Chiaravalloti et al. in "A Coding Support System for the ICD-9-CM Standard" are examples of systems developed for codification purposes [11]. The late one processes text in natural language using text mining algorithms, returning a list of possible codes for each case.

After the codification process, every discharge report can be read and perfectly understood in every country that adopts the same terminology.

For the purpose of the present work, and as stated by international directives, the ICD-9-CM codes are the basis of the DRGs decision [3]. Thereafter, the ICD-9-CM SPA coding tool was updated with the newer ICD-10-CM/PCS coding version, initially released in order to replace the ICD-9-CM version.

C. ICD-10-CM/PCS Clinical Coding

Firstly, it is important to note that ICD-10, unlike its precursor (ICD-9), is divided into two different parts, namely: [12], [13]

- ICD-10-CM for diagnosis coding;
- ICD-10-PCS for inpatient procedure coding.

When ICD-10 was defined, it was projected that ICD-10-CM would become a standard for all USA healthcare settings, whereas ICD-10-PCS would be required in inpatient settings only [12].

The ICD-10-CM/PCS clinical coding was mainly designed to offer notable advantages over ICD-9-CM, such as generating higher-quality clinical data. This would result in major improvements in the quality and the use of data in a significant number of healthcare settings, including driving a better healthcare management and improving significantly outcomes [12].

Supporters of ICD-10-CM/PCS defend that it incorporates greater specificity than its precursor in describing healthcare problems, but also clinical data, offers the addition of information relevant to ambulatory and managed care encounters, and impressively expanded injury codes that reflect the location of the injury [12], [14]–[16]. Thus, they praise its undeniable ability to provide a more detailed description of clinical situations, and substantially increasing the level of detail that can be captured [12], [13], [16], [17].

Furthermore, the structure of ICD-10-CM/PCS enables the possibility of greater expansion of code numbers, including risk factors that are regularly encountered in a primary care setting, instead of allowing the classification of diseases and injuries only [14].

In addition, while ICD-9 uses numeric codes (e.g., 001-999), ICD-10 uses an alphanumeric classification system which consists essentially of 1 letter followed by up to 3 numbers at a 4-character level (e.g., A00.0-Z99.9) [17], [18]. Thus, the alphanumeric format of ICD-10 offers a better structure than ICD-9, which allows a significant space for future revision without disruption of the numbering system [19].

On the other hand, the overall number of codes and diagnoses has increased significantly: about 17000 codes for ICD-9-CM to more than 155000 codes for ICD-10-CM/PCS [12], [13], [15]–[17], [20]. Thereby, the newer version includes previously unavailable codes since it enables a more detailed description of clinical situations, including for instance codes to distinguish between different types of diabetes or even the location on the patient's body of the health condition (e.g., left or right limb) [12].

Thus, while the main axis for ICD-9 is the nature of the health condition itself, the main axis of ICD-10 is the body region of the health condition with the highest level of specificity reachable [21].

In the long run, due to this improved level of detail, it is expected that ICD-10-CM will decrease medical fraud and abuse [12]. For instance, since even the location on the patient's body is registered in the codification process, it will reduce the possibility of coding professionals repeatedly reporting the same procedure on the same location of the body.

D. Interoperability: AIDA and AIDA-PCE

Nowadays, with the continuous growth of the clinical information stored into the hospital information systems (HISs), one of the major interests in the Medical Informatics field is to ensure interoperability between different information systems in health institutions [22], [23]. Thus, interoperability is increasingly considered a requirement in HISs rather than an option in order to implement an adequate communication and cooperation between distinct systems [24].

In short, the concept of interoperability can be defined as the ability of a system to communicate and share information with another system that arises in order to overcome the heterogeneity and distribution of several different sources of information. In the healthcare industry, the main goal of interoperability is to connect applications and data so that they can be shared across the organization and distributed to health professionals [22], [24].

Thereby, in this context, there is a need to implement dynamic platforms, such as multi-agent systems that allow the access and sharing of information between different information systems, in order to connect them, standardize distributed clinical systems, and thus reduce the delays normally generated in the process of sharing information [25].

The register of clinical information in CHP is ensured by a few HISs, namely the "*Sistema de Apoio ao Médico*" (SAM) – Medical Support System, the "*Sistema de Gestão de Doentes Hospitalares*" (SONHO) – Hospital Patient Management System, the "*Sistema de Apoio à Prática de Enfermagem*" (SAPE) – Nursing Support System, and the "*Processo Clínico Eletrónico*" (PCE) – Electronic Medical Record [23], [26].

In this context, the AIDA ("*Agência para Integração, Difusão e Arquivo de Informação Médica*") platform emerges. Some Portuguese health systems are equipped with the AIDA platform, including CHP, which uses proactive intelligent agents that ensure the interoperability between different and heterogeneous HISs, and other entities such as the complementary systems, including SAM, SONHO, SAPE, PCE, RIS ("Radiology Information System"), LIS ("Laboratory Information System"), DIS ("Department Information System"), and AIS ("Administrative Information System"), among others [22]–[26].

AIDA is a complex system consisting of simple and specialized subsystems, defined as intelligent agents, which are responsible for tasks such as the communication between heterogeneous systems, the sending and receiving of information (for example, clinical reports, medical images, and prescriptions), as well as the management and storage of data [22].

Thus, directly from AIDA it is possible to integrate, disseminate, and archive large sets of data from different sources (for example, services, departments, healthcare units, computers, and medical devices) [22]. In this way, the AIDA platform provides an easy access and sharing of registered information, facilitating medical research and the application and development of other computational tools, such as the ICD-9-CM and ICD-10-CM/PCS clinical coding tools described in this manuscript, in order to optimize the healthcare services provided by the health institution.

The electronic health record (EHR) is responsible for the safe and organized storage of all the information regarding a patient, from personal data to diagnoses and procedures [27]–[30].

The constant update is vital in this scenario, so the same research team from the Algoritmi Research Centre (University of Minho, Braga, Portugal) that developed AIDA put together the AIDA-PCE. Following the Problem Oriented Medical Record (POMR), all the patient information regarding symptoms, medical observations, diagnoses, and treatment plans are stored inside that structure.

Although these systems allow the insertion of free text and other non-universal information, the AIDA and AIDA-PCE present innovative and novel solutions to accomplish interoperability. On the other hand, the incorporation of the ICD-9-CM codes, and subsequently the ICD-10-CM/PCS codes, into the EHR represents an important feature in order to accomplish a cross-border medical record. The storage of ICD-9-CM and ICD-10-CM/PCS coded discharge reports saves space on the AIDA-PCE databases, and reduces medical errors.

The next section presents the main research methodologies followed to implement this work.

III. Research Methodologies

The realization of any study in the field of Information Technologies (ITs) includes the scrutinized research and analysis of the set of methodologies and technologies available and feasible in the design of the defined IT solutions. The choice of the most appropriate methods and tools is mostly based on the advantages pointed out, as well as on associated limitations and compliance issues with related systems.

Thus, the achievement of the SPA for ICD-9-CM, and its transition from the ICD-9-CM codification to the newer ICD-10-CM/PCS version, are based on the research methodology Design Science Research (DSR), mostly used in the construction and evaluation of useful and rigorous IT solutions. Each of the design phases presented in this study included the choice and use of the most appropriate methodologies, technologies, and tools for the definition and elaboration of the desired solution. Finally, a Proof of Concept (PoC) was also carried out corroborating the viability and usefulness of the clinical coding tool for ICD-9-CM designed and developed, and its update to the newer ICD-10-CM/PCS version, which consisted essentially of a SWOT analysis (Section V).

Thereby, a brief description of these two research methodologies are presented in this section, namely Design Science Research and Proof of Concept, in subsections A and B, respectively.

A. Design Science Research

In the area of ITs, the main objective of the use of the research methodology Design Science Research is the construction and evaluation of objects, also called "artefacts", that allow professionals to process organizational information and develop actions to solve a problem [31], [32].

Thus, the methodology that drove the realization of this project is the DSR. It consists of a rigorous method of scientific research used to develop successful artefacts [33]. It focuses on the IT artefact with a high priority in its relevancy in its application domain. Thus, in the context of solving real-world business problems, it is critical to try to improve the relevance and usefulness of the artefact [34], [35]. The designed appliance must correspond to a viable technological solution for solving important and relevant business problems, and its usefulness, quality, and effectiveness must be rigorously demonstrated through well-executed evaluation methods. In addition, research should provide clear and verifiable contributions, and should be based on the application of rigorous methods in its construction and evaluation process [34], [35].

In Figure 1, the research methodology DSR is outlined, that is, its different interconnected steps that synthesize the steps to be followed through the DSR in the construction of scientific IT artefacts, namely the steps of “Identify Problem & Motivate”, “Define Objectives of a Solution”, “Design & Development”, “Demonstration”, “Evaluation”, and “Communication”. These are the phases adopted in the design of this case study.

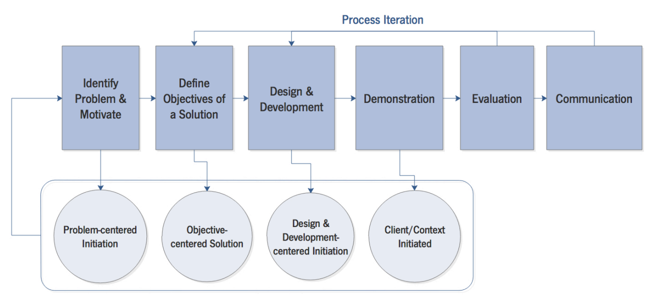


Figure 1. Schematic representation of the research methodology Design Science Research (adapted from [33]).

In short, in the first steps, the problem and the motivation are defined, as well as the objectives of the solution found. Then the artefact is designed and developed, directed to an important business problem to be solved that must be relevant to the solution of the same. Its development must follow a rigorous scientific process based on the knowledge and the theory already explored. Finally, the solution must be demonstrated, evaluated, communicated, and propagated efficiently to the target audience [31], [36].

Thus, the case study described in this manuscript follows the DSR research methodology because the IT solution defined meets the needs of health professionals of CHP, that is, a new clinical coding tool to support clinical practice assisting in their work, meeting the challenges currently existing at CHP's facilities.

Therefore, it provides the health institution with an appropriate and well-founded solution, based on methods and technologies that have already been explored and adapted to solve the problem in question, and also stimulate new knowledge for the organization and the scientific community. Thus, the development of this project additionally included the dissemination of the IT artifact to CHP's professionals, as well as the writing of scientific papers.

Finally, it is important to note that the clinical tool developed, including its update to the ICD-10-CM/PCS version, was duly evaluated through a SWOT analysis. It should also be pointed out that the application of the Proof of Concept research methodology to prove the feasibility, usefulness, and usability of the tool, which included a SWOT analysis, is briefly described in Subsection B.

B. *Proof of Concept*

The Proof of Concept research methodology consists of a practical model that can prove or validate the concept established through analysis or even technical articles. Hence, it goes on to verify whether a concept or theory is successful and feasible and, on the other hand, is thus susceptible of being exploited in a useful way [37].

Therefore, a PoC is often pointed out as one of the most important steps in the design, development, implementation, and proposal process of a prototype of a IT solution, mainly to establish if an IT solution fulfills its purpose, that is, it meets the requirements and defined objectives for which it was originally designed. On the other hand, it also allows the identification of potential failures or errors in the IT solution developed [38].

Summarizing, a PoC allows to demonstrate in practice the concepts, methodologies, and technologies involved in the elaboration of a given project and thus validate the proposed solution by proving its feasibility and usefulness for the purpose for which it is intended by defending its potential.

In this study, the defense of the viability and usefulness of the clinical coding tool for ICD-9-CM went through the application of the Proof of Concept research methodology, including its subsequent transition to the newer ICD-10-CM/PCS coding version. Thus, a SWOT analysis was carried out for the Web application.

The proof of concept of this project is briefly described in Section V of this article.

The next section presents the SPA developed and implemented for the ICD-9-CM coding version, as well as its transition to ICD-10-CM/PCS.

IV. Single-Page Application

The codification of the discharge reports was made manually in CHP, making the process too slow and with a high error probability. Therefore, it emerged the need to create a process that would reduce the codification time. Thus, it was developed a SPA through which the health professionals are able to perform the codification process, and at the same time to consult patients' data, such as the discharge report, the personal information, and the hospital services where the patient was admitted.

The first version was created based on the ICD-9-CM coding version. However, there was a need to adopt a more advanced coding system, namely the ICD-10-CM/PCS codification.

A. *SPA Version 1*

The purposed layout in this scenario encompasses a solution with three main components: the patient information, the codification area, and the discharge report. The codification area is divided into five frames: diagnoses, external causes, procedures, tumor morphology, and observations. All the boards, with exception of the observations board, which is the only one that allows free text insertion, are composed by rows divided mainly in priority, description, and code.

The dynamic and aided search leads to a faster process than the already existing one. When a word is typed in the description field, a list of all the ICD-9-CM codes is presented, and when the user picks one of them on the description, the respective code is automatically filled. On the other hand, when the user enters the code, the description is also automatically filled.

When a codification is finished, the user sends it to an evaluator. If the codification fails in this evaluation phase for any reason, it is sent again to the list of codifications that needs to be done. Thus, the application was developed with three different modules. One to be used in the first codification, another to be used when a discharge report was already codified and for some reason failed in the evaluation phase, and one final module to be used when the user only has view permissions. In this last module, the users are not able to change any field present in the codification.

The application development leaned on the LAMP architecture. It uses Linux as the operative system, Apache as the Web server, MySQL as the relational database management system (RDBMS), and PHP as the object-oriented language.

In order to develop a fluid and dynamic application, the AngularJS framework greatly contributed. The modularity and extensibility of this JavaScript framework allows the development of diverse and futuristic applications.

The database storages all the data related to the codification codes, the discharge reports, user information, and all the information generated by the codification process. The RESTful Web service mediates the communication between the SPA and the database.

B. *SPA Version 2*

In the second version of the SPA, all the SPA architecture remained the same. The difference from the first version to the second was only the use of a different codification system, which consequently forced a layout change in the codification area. In this version, the codification area only presents three frames: diagnoses, procedures, and observations.

The dynamic and aided search is still present in this version.

Once the ICD-10-CM/PCS codification presents a hierarchical structure, it is easier to find a code since the order is followed, so, it was added the help button in the code insertion. When the user selects this button, it opens a new frame with dynamic search. This frame is divided into seven parts: section, body system, root operation, body part, approach, device, and qualifier. At first, all components are empty except for the component section. Since this is the root of coding, it represents the first character of the code. When the user selects a section, the body system is automatically filled with options that are related with the section selected. In turn, when the body system is selected, it automatically fills the root operation options. When the three components section, body system, and root operation are selected, a board with all the other components is filled. And after all seven parts have been selected, the code is finally complete and, thereafter, it is sent to the main frame to the row where the user clicked in the help button. It should additionally be noted that a process of validation for the code exists.

Although this second version also use MySQL as the RDBMS, the tables of the database needed to be changed in order to be able to store the new coding system.

The next section presents the SWOT analysis of the Web application developed.

V. SWOT ANALYSIS

To test the viability, the utility, the quality, and the efficiency of the application, a PoC was necessary, in this case, a SWOT analysis. This analysis allows to analysis the strengths, weaknesses, opportunities, and threats of the application [39].

With the update of the coding system for the ICD-10-CM/PCS codification, a weakness of the project was eliminated, thus, becoming a more up-to-date project. Therefore, it was made a SWOT analysis of the second version of the project.

Strengths:

- High usability, intuitive, and easy to learn (user-friendly);
- Easy access to the data of patients, as well as the hospital services in which the patient was;
- High scalability;
- Easy of reissue of coded discharge reports;
- Decrease of the codification time of the discharge reports;
- Decrease of human error;
- Easy adaptability to different health institutions.

Weaknesses:

- Requires internet connection.

Opportunities:

- Modernization and organizational development;
- Increasing expectation of the hospital administration to obtain methods that facilitate the hospital financing calculation;
- Provide the tool to help in the calculation of the hospital financing.

Threats:

- Lack of acceptance to resort to new technologies by health professionals.

The next section presents the conclusion and future work of this study.

VI. CONCLUSION AND FUTURE WORK

Finally, the realization of this case study allowed the development of a clinical practice tool, namely a user-friendly clinical coding tool for ICD-9-CM, and its subsequent transition to the newer ICD-10-CM/PCS coding version. The Web application is currently implemented in a production machine of CHP, and it is currently being used by the coding professionals of the hospital in order to perform the clinical coding of the episodes of hospital discharges from patients admitted to CHP. This will then facilitate the grouping of processes into DRGs, that is, a financial system that can manage the costs and waste associated with healthcare services. In the coming years, the expansion of the Web application is expected.

Trained professionals using the ICD-9-CM and ICD-10-CM/PCS clinical coding tools reported significant differences in time consumption and committed errors when using a computerized system to perform their tasks. It represents an asset to its users, since it facilitates the work of health professionals, and increases their capacity and speed of work by reducing the number of tasks required to perform a certain codification. In this way, the development of the clinical tool allows the centralization of a set of tasks and information in a SPA, greatly benefiting its users.

When comparing the SPA for ICD-9-CM codification with its update to the newer ICD-10-CM/PCS version, health professionals defend that the advantages are much greater, including a better specificity in describing clinical situations, leading to an improved level of detail. They also defend that it is easier to find a code in the new update due to the help bottom. Thus, they do not need to remember all the codes or their entire description, making it more intuitive.

Regarding future work, the addition of a Business Intelligence (BI) module in the clinical coding tool for ICD-10-CM/PCS is foreseen, that is, the addition of a module with clinical and performance indicators [40]. Its principal aim is the visualization of indicators that show the association between the number of coded processes and each coding professional, as well as the temporal evolution of the number of processes encoded by each coding professional. Thus, the main objective of the insertion of this module is to study and analyze the performance of the coding professionals, that is, to identify, for example, the coding professionals who codify the most, and also those who are coding the least. In this way, it is tried to encourage even more the increase of the production of health professionals at CHP.

Finally, in early 2018, the research team already began the implementation of the system in more health institutions in Portugal. It is thus confirmed our strong desire to continue to expand the system across the country due to its undeniable advantages, which were defended throughout this paper.

ACKNOWLEDGEMENT

This work has been supported by Compete POCI-01-0145-FEDER-007043 and FCT - *Fundação para a Ciência e Tecnologia* within the Project Scope UID/CEC/00319/2013.

REFERENCES

[1] C. Coimbra *et al.*, "Improving Diagnosis-related Groups with a Computerized Clinical Coding Tool for ICD-9-CM Codification," in *eTELEMED 2017: The Ninth International Conference on eHealth, Telemedicine, and Social Medicine*, 2017, pp. 33–36.

[2] M. M. Bellanger, W. Quentin, and S. S. Tan, "Childbirth and Diagnosis Related Groups (DRGs): Patient Classification and Hospital Reimbursement in 11 European Countries," *Eur. J. Obstet. Gynecol. Reprod. Biol.*, vol. 168, no. 1, pp. 12–19, May 2013.

[3] W. Quentin, D. Scheller-Kreinsen, M. Blumel, A. Geissler, and R. Busse, "Hospital Payment Based On Diagnosis-Related Groups Differs In Europe And Holds Lessons For The United States," *Health Aff.*, vol. 32, no. 4, pp. 713–723, Apr. 2013.

[4] Y. Xin, "Comparison of Hospital Medical Waste Generation Rate Based on Diagnosis-related Groups," *J. Clean. Prod.*, vol. 100, pp. 202–207, 2015.

[5] M. Barrento, "Sistema de Business Intelligence Aplicado aos Grupos de Diagnósticos Homogéneos," Universidade Nova de Lisboa, 2010.

[6] C. Fourie, N. Biller-Andorno, and V. Wild, "Systematically Evaluating the Impact of Diagnosis-related Groups (DRGs) on Healthcare Delivery: A Matrix of Ethical Implications," *Health Policy (New. York).*, vol. 115, no. 2–3, pp. 157–164, 2014.

[7] R. Busse *et al.*, "Diagnosis Related Groups in Europe: Moving Towards Transparency, Efficiency, and Quality in Hospitals?," *BMJ*, vol. 346, 2013.

[8] P. Jorge, "Plataforma de Dados de Saúde - Portal Institucional," Instituto Superior de Engenharia do Porto, 2013.

[9] S. B. Rimler, B. D. Gale, and D. L. Reede, "Diagnosis-related Groups and Hospital Inpatient Federal Reimbursement," *RadioGraphics*, vol. 35, no. 6, pp. 1825–1834, Oct. 2015.

[10] M. Rouse and D. Fluckinger, "ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification)," 2014. [Online]. Available: http://searchhealthit.techtarget.com/definition/ICD-9-CM. [Accessed: 31-Jan-2017].

[11] M. T. Chiaravalloti, R. Guarasci, V. Lagani, E. Pasceri, and R. Trunfio, "A Coding Support System for the ICD-9-CM Standard," *Proc. - 2014 IEEE Int. Conf. Healthc. Informatics, ICHI 2014*, pp. 71–78, 2014.

[12] M. Topaz, L. Shafran-Topaz, and K. H. Bowles, "ICD-9 to ICD-10: Evolution, Revolution, and Current Debates in the United States," *Perspect. Heal. Inf. Manag. / AHIMA (American Heal. Inf. Manag. Assoc.*, vol. 10, no. Spring, 2013.

[13] R. Mills, R. Butler, E. McCullough, M. Bao, and R. Averill, "Impact of the Transition to ICD-10 on Medicare Inpatient Hospital Payments," *Medicare Medicaid Res. Rev.*, vol. 2, no. 2, pp. 1–13, 2011.

[14] A. Hazlewood, "ICD-9 CM to ICD-10 CM: Implementation Issues and Challenges," *The American Health Information Management Association*, 2017. [Online]. Available: http://library.ahima.org/doc?oid=59978#.WUqLXMZOrVo. [Accessed: 21-Jun-2017].

[15] J. A. Hirsch *et al.*, "ICD-10: History and Context," *Am. J. Neuroradiol.*, vol. 37, no. 4, pp. 596–599, 2016.

[16] A. G. Fiks and R. W. Grundmeier, "Elucidating Challenges and Opportunities in the Transition to ICD-10-CM," *Pediatrics*, vol. 134, no. 1, pp. 169–170, 2014.

[17] J. Armstrong, P. McDermott, G. R. Saade, and S. K. Srinivas, "Coding Update of the SMFM Definition of Low Risk Cesarean Delivery from ICD-9-CM to ICD-10-CM," *Am. J. Obstet. Gynecol.*, 2017.

[18] N. Jetté *et al.*, "The Development, Evolution, and Modifications of ICD-10," *Med. Care*, vol. 48, no. 12, pp. 1105–1110, 2010.

[19] S. Bowman, "Why ICD-10 Is Worth the Trouble," *The American Health Information Management Association*, 2017. [Online]. Available: http://library.ahima.org/doc?oid=78673#.WUqMIMZOrVo. [Accessed: 21-Jun-2017].

[20] R. Caskey *et al.*, "The Transition to ICD-10-CM: Challenges for Pediatric Practice," *Pediatrics*, vol. 134, no. 1, pp. 31–36, 2014.

[21] J. R. Lowe, G. Raugi, G. E. Reiber, and J. D. Whitney, "Changes in Classifications of Chronic Lower - Limb Wound Codes in Patients with Diabetes: ICD-9-CM Versus ICD-10-CM," *Adv. Skin Wound Care*, vol. 28, no. 2, pp. 84–92, 2015.

[22] L. Cardoso, F. Marins, F. Portela, A. Abelha, and J. Machado, "Healthcare Interoperability through Intelligent Agent Technology," *Procedia Technol.*, vol. 16, pp. 1334–1341, 2014.

[23] H. Peixoto, M. Santos, A. Abelha, and J. Machado, "Intelligence in Interoperability with AIDA," in *Foundations of Intelligent Systems*, Springer Berlin Heidelberg, 2012, pp. 264–273.

[24] L. Cardoso, F. Marins, F. Portela, M. Santos, A. Abelha, and J. Machado, "The Next Generation of Interoperability Agents in Healthcare," *Int. J. Environ.*

Res. Public Health, vol. 11, no. 5, pp. 5349–5371, 2014.

[25] A. Abelha, C. Analide, J. Machado, J. Neves, M. Santos, and P. Novais, “Ambient Intelligence and Simulation in Health Care Virtual Scenarios,” *IFIP Int. Fed. Inf. Process.*, vol. 243, pp. 461–468, 2007.

[26] J. Machado, A. Abelha, J. Neves, and M. Santos, “Ambient Intelligence in Medicine,” in *IEEE Biomedical Circuits and Systems Conference*, 2006, pp. 94–97.

[27] M. Salazar *et al.*, “Step towards Paper Free Hospital through Electronic Health Record,” in *Advances in Information Systems and Technologies*, Springer International Publishing, 2013, pp. 685–694.

[28] J. Duarte *et al.*, “Data Quality Evaluation of Electronic Health Records in the Hospital Admission Process,” in *EEE/ACIS 9th International Conference on Computer and Information Science*, 2010, pp. 201–206.

[29] J. Duarte, C. F. Portela, A. Abelha, J. Machado, and M. F. Santos, “Electronic Health Record in Dermatology Service,” in *ENTERprise Information Systems*, Springer Berlin Heidelberg, 2011, pp. 156–164.

[30] F. Portela *et al.*, “Electronic Health Records in the Emergency Room,” in *IEEE/ACIS 9th International Conference on Computer and Information Science*, 2010, pp. 195–200.

[31] S. T. March and V. C. Storey, “Design Science in the Information Systems discipline: An introduction to the special issue on Design Science Research,” *MIS Quartly*, vol. 32, no. 4, pp. 725–730, 2008.

[32] M. Bilandzic and J. Venable, “Towards Participatory Action Design Research: Adapting Action Research and Design Science Research Methods for Urban Informatics,” *J. Community Informatics*, vol. 7, no. 3, 2011.

[33] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A Design Science Research Methodology for Information Systems Research,” *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, 2007.

[34] a. R. Hevner, S. T. March, and J. Park, “Design Science in Information Systems Research,” *MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004.

[35] A. R. Hevner and S. Chatterjee, “Design Research in Information Systems,” *Des. Res. Inf. Syst.*, vol. 28, pp. 75–105, 2004.

[36] V. K. Vaishnavi and W. Kuechler, *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, 2nd ed. CRC Press, 2015.

[37] A. B. Sergey, D. B. Alexandr, and A. T. Sergey, “Proof of Concept Center — A Promising Tool for Innovative Development at Entrepreneurial Universities,” in *Procedia - Social and Behavioral Sciences*, 2015, vol. 166, pp. 240–245.

[38] B. Schmidt, “Proof of Principle Studies,” *Epilepsy Res.*, vol. 68, no. 1, pp. 48–52, 2006.

[39] R. Pereira, M. Salazar, A. Abelha, and J. Machado, “SWOT Analysis of a Portuguese Electronic Health Record,” in *Collaborative, Trusted and Privacy-Aware e/m-Services*, Springer Berlin Heidelberg, 2013, pp. 169–177.

[40] A. Brandão *et al.*, “A Benchmarking Analysis of Open-Source Business Intelligence Tools in Healthcare Environments,” *Information*, vol. 7, no. 4, p. 57, Oct. 2016.

Monitoring of Non-stationary Health-Recovery Processes with Control Charts

Olgierd Hryniewicz and Katarzyna Kaczmarek-Majer

Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
Email: {hryniewi,K.Kaczmarek}@ibspan.waw.pl

***Abstract*—The paper presents a statistical process control method for monitoring health-recovery processes described by short non-stationary time series. The Shewhart control chart for residuals, based on model averaging approach, is built for differences between values of consecutive observations. The practical applicability of this new approach has been demonstrated using a real-life example of a recovery from a mild hypertension episode.**

Keywords–E-health; Control chart for residuals; Short time series; Non-stationary process; Stability of recovery process.

I. Introduction

This paper is a significant extension of the conference paper "Monitoring of Health-Recovery Processes with Control Charts" presented in the Proceedings of ACCSE'2017 conference, held in Venice, Italy [1], and is focused on monitoring of the stability of non-stationary processes. Stability is an important feature of many processes. A process is considered stable, or under control, when its uncontrolled variation is purely random (e.g., due to random measurement errors). In 1924 W. Shewhart introduced a simple tool for monitoring stable processes - a control chart. In its initial stage, which is assumed to be in-control, monitored process characteristics are measured, and their mean value and standard deviation are recorded. These recorded values are used for the design of a control chart, known as the Shewhart control chart, which consists of control lines: central, and two (or one, when only deviation of a process level in one direction is interesting) control. The central line represents the mean value of the process level (or a certain target value for this process), and control lines are located at three standard deviations from a central line. The process is considered stable when its future observations are located inside control lines (limits). When an observation falls outside the control lines, an alarm signal is generated, and the process is considered as being possibly out of control (unstable). When a monitored process goes out of control, it is recommended to look for the reason of this, and take appropriate actions with the aim to revert it to the in-control state.

Within this paper, we discuss the construction of a control chart for autocorrelated and non-stationary processes. In the experimental part, we extend the results presented in [1], and related to a special kind of medical data, namely data describing health-recovery processes.

A. Monitoring of health-recovery processes

For many years, physicians have been prescribing certain treatments, and advances in the health recovery of a treated patient have been monitored during visits, e.g., in health care units. Therefore, possible failures of applied treatments were usually disclosed with delay. In many cases, such delays have had detrimental effects on patient's health. However, with the development of e-health systems based on telemedicine this situation has been dramatically changed. Nowadays, it is possible to monitor the state of patient's health even continuously. However, the main problem now is not related to measurements and transmission of data, but to processing of available information. When human's life is endangered, very expensive systems, e.g., in intensive care hospital units, are used. However, in many cases, the usage of all those sophisticated Information Technology (IT) systems is not necessary. It is quite sufficient to process data off-line, and to signal only these cases when consultancy or intervention of a physician is really needed. What is important in this context, it is the stability of health-recovery processes, understood as non-existence of abnormal and unpredictable changes of the monitored process. It has to be noted here, that an unstable process may be still inside some "normal limits", pre-established by physicians, but its revealed instability suggests the possibility of going beyond such limits. Monitoring of such stability can be achieved by the usage of appropriately designed control charts. The proposal of such monitoring processes, based on a control chart for so called residuals, is the main purpose of this paper.

It has to be noted here that the design of monitoring procedures, such as these related to health care, has to fulfill two conflicting requirements. A procedure has to be designed using an appropriate mathematical model of a monitored process, but on the other hand, it has to be simple to use, and the results of this procedure have to be easily interpreted, even by users without any special statistical training. This conflict is especially visible in monitoring health-related processes. For example, processes continuously monitored in intensive care units, or processes monitored by wearable sensors, are very complex, because their mathematical models have to take into account many different circumstances, such like natural 24-hours variability or drug administration. Unfortunately, monitoring procedures based on such complex mathematical models, if not fully automated, are very difficult to interpret. In this paper we consider a particular case when a mathematical model is still relatively simple, but the results of its application are much easier for interpretation. By taking this assumption we restrict applicability of the proposed procedure to cases, in which measurements are made in comparable circumstances, e.g., once a day at the same time.

B. Real-life example of a health-recovery process

The experimental part of this paper is a case study. Daily recordings of blood pressure (BP) for a period of 480 days of a patient who is under treatment against mild blood hypertension are analyzed and discussed. In Figure 1, we present the considered process of one-a-day BP measurements.

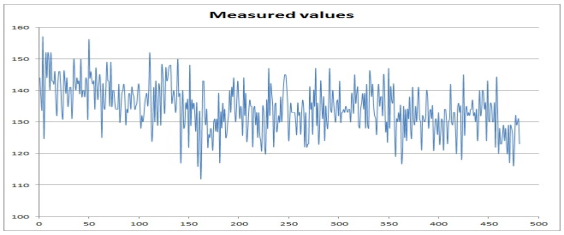


Figure 1. Measurements of blood pressure.

In the case of health-related time series data underlying processes are often non-stationary and correlated, and therefore, proper establishment of control limits is challenging. This paper investigates this problem.

C. Related work

Basic control charts, used in over 90% practical applications, are designed under two main assumptions: statistical independence of consecutive observations, and the normal distribution of measured characteristics. However, in many practical cases, especially when individual process observations are monitored, these assumptions are not fulfilled. Thus, in the recent 40 years, many inspection procedures that do not rely on these assumptions have been proposed. They are usually described in scientific journal papers or in a few textbooks on statistical quality control, such as a famous book of Montgomery [2]. They have been applied in many areas, but only some of them have been applied in health-related services, and similar applications, where their applicability seems to be quite natural. One of these natural applications, monitoring of health-recovery processes, has been described in [1]. A comprehensive review of other applications of control charts in health-care and public-health surveillance, written from a perspective of a quality control specialist, can be found in the paper by Woodall [3]. Similar review papers, but written from a perspective of a health care specialist, were written by Tennant et al. [4], Thor et al. [5], and Winkel and Zhang [6]. Since the time of the publication of these papers, many other papers on this topic have been published, mainly in journals related to medicine.

Despite real popularity of control charts in many areas, such as industry, finance and business, the number of their applications in health care is relatively small. Probably the main reason for this situation was given by the authors of [4] who wrote in their conclusions “Control charts appear to have a promising but largely under-researched role in monitoring clinical variables in individual patients”. However, the situation has been changed during the last few years. In their editorial titled “Statistical process control in healthcare improvement - new kid on the block?”, and published in a recent issue of Acta Anaesthesiologica Scandinavica, Møller and Anhøj [7] have written that the yearly number of PubMed citations including the term “statistical process control” in any field has increased from a level of 10 in 2010 to the level of 90 in 2018. If we take a perspective of a statistician, a probable reason of this situation is incompatibility of basic assumptions used for construction of statistical process control (SPC) tools, such as control charts, and the reality of health care. For example, consecutive observations of health-related characteristics are seldom independent. Moreover, they are often described by non-stationary random processes, and the runs of interesting observations are short. Therefore, control charts described in popular textbooks, and in the great majority of scientific papers, are not appropriate for monitoring such processes, and the results of using such charts may be unsatisfactory. Some new, more appropriate, approaches have been investigated quite recently. For example, the properties of control charts used for short runs for autocorrelated, but stationary, data have been discussed in [8].

First research papers on applications of statistical process control in cardiology have been published already in the 1990’s. For example, G. Cornélissen et al. [9] use control charts to monitor blood pressure and heart rate for individualized assessment of a patient’s response to a drug. Another application of control charts in monitoring of blood pressure has been proposed recently in the paper by Albloushi et al. [10]. Furthermore, benefits of applications of control charts in cardiology are constantly being confirmed by several papers on this topic published each year. For example, see the recent paper of Jung and Kim [11], who introduced an electrocardiographic (ECG) monitoring procedure for diagnosing PVC beats using a wavelet-based statistical process control methodology, or the paper of Lambeth et al. [12], who adapted the statistical process control methods to monitor the stability of admission temperature and glucose-level processes for the very low birth-weight infants within first hour of birth at a neonatal intensive care unit (NICU).

The paper is organized as follows. In Section II, we describe a mathematical model of a stochastic process (a time series) that may be useful for the description of health-recovery data. Then, in Section III, we propose a control chart based procedure that may be used for monitoring non-stationary health-recovery processes. The problem of the monitoring of short time series using the sXWAM chart, proposed by us in [13], is considered in Section IV. In Section V, results for the real-life health-recovery process are presented and discussed. The paper is concluded in its last section, where we discuss limitations of the proposed methodology, and indicate some possible future applications.

II. Mathematical model of a monitored process

Series of dependent observations may be described by many mathematical models. When the expected value (the mean), the variance, and the covariances of the underlying process are constant in time, we call such processes stationary. When these statistical characteristics vary in time (e.g., according to a certain trend function) we call such processes non-stationary. Statistical methods of the analysis of time series, both stationary and non-stationary, can be found in many textbooks, such as, e.g., the book by Brockwell and Davis [14].

Simple visual analysis of the considered data (Fig. 1) reveals the existence of a visible trend. Thus, this time series cannot be described by a simple mathematical model. There are formal ways, namely statistical tests, to verify stationarity of a time series. One of the most powerful tests, the Kwiatkowski-Philips-Schmidt-Shin (KPSS) test, was introduced by Kwiatkowski et al. [15], and enables testing the null hypothesis of stationarity, either around a level or around a linear trend, against the alternative of a unit root. For the considered exemplary time series, the value of the KPSS statistic amounts to 2.9737 and its related p-value is smaller than 0.01. Basing on this result, we reject the null hypothesis of stationarity, concluding that the considered time series is non-stationary.

However, there exists a large class of non-stationary processes that can be transformed to stationary processes, which are much easier for analysis. An important member of this class of non-stationary time series can be described by the Autoregressive Integrated Moving Average (ARIMA) model, introduced in the seminal book by Box and Jenkins [16]. For an ARIMA non-stationary process of first order, differences between consecutive observations are described by a stationary Autoregressive Moving Average (ARMA) process. The time series $\{X_t\}$ is ARMA(p,q) process if $\{X_t\}$ is stationary and if for every t [14],

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{1}$$

where $\{Z_t\}$ is a purely random (white noise) process of uncorrelated observations having null expected values and the same finite variances, and the polynomials $(1-\phi_1 z - \cdots - \phi_p z^p)$ and $(1+\theta_1 z + \cdots + \theta_q z^q)$ have no common factors. The special case of the ARMA(p,q) model is the autoregression model AR(p) defined by the following equation

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t. \tag{2}$$

When the process $\{Z_t\}$ has a certain non-zero, and constant in time, expected value μ, the model described by (2) has to be slightly modified, and in this case has the following form

$$X_t - \mu - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t. \tag{3}$$

Now, let us look at Figure 2, where differences between the values of consecutive observations of the process presented in Figure 1 are displayed.

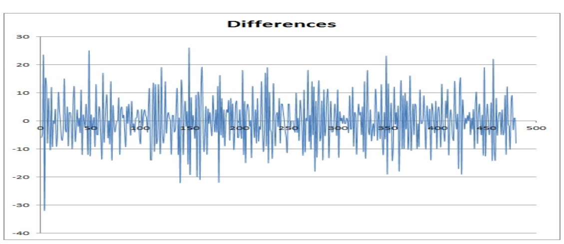


Figure 2. Differences between consecutive measurements.

Basing on visual inspection, the process displayed in Figure 2 looks stationary. The statistical tests also confirm its stationarity. The KPSS statistic is 0.0107, and the p-value amounts to 0.1, thus there are no reasons to reject the null hypothesis of stationarity. Hence, we can conclude that this time series is stationary, and may be well described by an autoregression process of the fourth order $AR(-0.862, -0.713, -0.358, -0.207)$. The constant term μ in this case is close to zero, and equal to -0.085. Therefore, this real-life example has motivated us to consider in this paper time series described by models of a similar type.

Let $X_1, X_2, \ldots, X_n$ be a series of measurements obtained during a period of time when a monitored process may be considered (e.g., according to a physician who supervises the treatment) as stable. The process of first differences is now defined as follows: $D_i = X_{i+1} - X_i, i = 1, \ldots, n-1$. We assume the ith difference is related to the previous observations according to the equation

$$D_i = \mu + a_1 * d_{i-1} + a_2 * d_{i-2} + \cdots + a_p * d_{i-p} + \epsilon_i, i = p+1, \ldots, \tag{4}$$

where $\epsilon_i, i = p+1, \ldots$ are normally distributed independent random variables with the expected value equal to zero, and the same finite standard deviation.

Estimation of the model $AR(p)$, given by (4), is relatively simple when we know the order of the model p. In order to find estimators $\hat{a}_1, \ldots, \hat{a}_p$, we have to calculate first p sample autocorrelations $r_1, r_2, \ldots, r_p$, defined as

$$r_i = \frac{n \sum_{t=1}^{n-i} (d_t - \hat{\mu})(d_{t+i} - \hat{\mu})}{(n-i) \sum_{t=1}^{n} (d_t - \hat{\mu})^2}, i = 1, \ldots, p, \tag{5}$$

where n is the number of observations in the sample (usually, it is assumed that $n \geq 4p$), and $\hat{\mu}$ is the sample average. Then, the parameters $a_1, \ldots, a_p$ of the $AR(p)$ model are calculated by solving the Yule-Walker equations (see, [14])

$$\begin{array}{l} r_1 = a_1 + a_2 r_1 + \ldots + a_p r_{p-1} \\ r_2 = a_1 r_1 + a_2 + \ldots + a_p r_{p-2} \\ \ldots \\ r_p = a_1 r_{p-1} + a_2 r_{p-2} + \ldots + a_p \end{array} \tag{6}$$

When a more general model of the AR process (3) has to be used, the constant term μ may be estimated using the following equation

$$\hat{\mu} = \bar{x}(1 - a_1 - \cdots - a_p), \tag{7}$$

where $\bar{x}$ denotes the average of the observed process values.

The estimators obtained by solving the Yule-Walker equations are, unfortunately, not numerically stable, especially for small sample sizes. A better method was proposed by Burg. A good description of Burg's algorithm can be found in [17]. Burg's algorithm is used to solve the following optimization problem: for the set of observations $x_1, \ldots, x_N$ find the values $a_1^*, \ldots, a_k^*$ that minimize F_k defined as

$$F_k = \sum_{n=k}^{N} (x_n - (-\sum_{i=1}^{k} a_i x_{n-i}))^2 \tag{8}$$

The estimators of the $AR(p)$ model given by (2) are obtained by setting $k = p, N = n$, $x_i = d_i, i = 1, \ldots, n-1$, and $\hat{a}_i = -a_i^*, i = 1, \ldots, p$. Note, that in this formulation of the optimization problem we do not use a constant term μ. Therefore, the models estimated using the Yule-Walker equations (moment estimators), and the models estimated using Burg's algorithm, may be different.

In both presented methods for the estimation of the AR model we have assumed that the model order p is known in advance. In practice, however, we do not know the order of the autoregression process, so we need to estimate p from data. In order to do this, we define a transformed random variable, called the *residual*. In the case of autoregression processes, considered in this paper, the residual is defined as

$$Z_i = D_i - (\hat{\mu} + a_1 d_{i-1} + \ldots + a_p d_{i-p}), i = p+1, \ldots, n. \quad (9)$$

When we know exactly the autoregression model, the probability distribution of residuals is the same as the distribution of random variables $\epsilon_i, i = 1, \ldots$ in (2), and its variance can be used as a measure of the accuracy of predictions of future values of the process. For given sample data of size n, the variance of residuals is decreasing with the increasing values of p. However, the estimates of p model's parameters $a_1, \ldots, a_p$ become less precise, and thus the overall precision of prediction with future data deteriorates. As the remedy to this effect, several optimization criteria with a penalty factor, which discourages the fitting of models with too many parameters, have been proposed. In this research we have used the criterion proposed by Akaike [18], and defined as

$$\begin{aligned} BIC = \; & (n-p)\ln[n\hat{\sigma}^2/(n-p)] + n(1+\ln\sqrt{2\pi}) + \\ & p\ln[(\textstyle\sum_{t=1}^{n} d_t^2 - n\hat{\sigma}^2)/p], \end{aligned} \quad (10)$$

where d_t are our transformed process observations centered in such a way that their expected values are equal to zero, and $\hat{\sigma}^2$ is the observed variance of residuals. The fitted model, i.e., the estimated order p and parameters of the model $\hat{a}_1, \ldots, \hat{a}_p$ (and $\hat{\mu}$, if a constant term is used) minimizes the value of BIC calculated according to (10). We will use this model in the construction of a control chart for monitoring health-recovery processes.

III. CONTROL CHART FOR PROCESS MONITORING WITH AUTOCORRELATED DATA

A. Design of a chart

The design of a simple Shewhart control chart, in the case of a sufficiently large number of individual and mutually independent observations, is extremely simple. One has to collect data (a sample) from a period when the monitored process is stable, calculate average value $\bar{x}$ and standard deviation σ_x, and set the control limits, upper (CUP) and lower (CLO), to

$$\begin{aligned} CUP &= \bar{x} + 3\sigma_x \\ CLO &= \bar{x} - 3\sigma_x. \end{aligned} \quad (11)$$

When process deterioration is related only to increase (decrease) of a process level, one can use one-sided control charts with respective upper (lower) control limits. Usually, it is assumed that the monitored characteristic is normally distributed, and in this case the probability of observing the observation outside one control limit when the monitored process is stable (i.e., observing a false alarm) is very low, and equals 0.00135. It means, that the expected number of observations between consecutive false alarms is equal approximately to 740 (for a one-sided chart), or to 370 (for a two-sided chart).

When consecutive observations of a monitored process are statistically dependent, the situation becomes much more complicated. For example, when sample data are autocorrelated, the properties of a control chart designed using a simple algorithm described above may be completely different from those observed for independent data. Consider, for example, a Shewhart control chart constructed using the results of first 20 measurements, and presented in Figure 3.

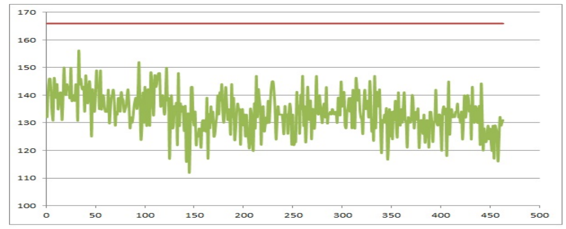


Figure 3. Control chart for original measurements.

In this case we use the chart with only an upper control limit because alarms should be triggered only by abrupt increases of blood tension. If we look at Figure 3 we may immediately notice that the charted values of measurements are rather far from the upper control limit. Moreover, the difference between the observed process level and the control limit increases in time. Therefore, the ability of the chart to trigger an alarm deteriorates in time, and for this reason the chart constructed using a classical way becomes useless.

The problem presented in Figure 3 is frequently observed when charted data are autocorrelated, as it is the case of our measurements. To cope with this problem, statisticians have proposed two general approaches. In the first one, we chart the original data, but control limits are adjusted using the knowledge about the type of dependence. In the second general approach, originally introduced by Alwan and Roberts [19], a control chart is used for monitoring residuals. Their methodology is applicable for any class of processes, so it is also applicable for the autoregression process of differences D_i considered in this paper. According to the methodology proposed by Alwan and Roberts [19], the deterministic part of (2) or (3) is estimated from sample data of n elements, and used for the calculation of residuals according to (9). Then, these residuals are used for the construction of our control chart according to the algorithm described above.

It is worth noticing that the Shewhart control chart for individual observations, also known as the X chart, is not the only control chart used for monitoring stability of monitored processes. However, it is the simplest one. Moreover, it is easy to interpret by non-specialists. This second feature seems to us very important if we have to use it in a simple health-care monitoring procedure.

B. Operating procedure

Operating procedure of the proposed control chart for residuals, applied for differences between consecutive observations of the monitored process, is the same as in the case of a classical Shewhart control chart. Using the estimated process model, we calculate the predicted value of the difference between the next two observations of the monitored process. An alarm signal is generated when an observed residual (difference between an observed and predicted values) falls beyond control lines. In Figure 5, we present a one-sided (with an upper control limit) control chart for residuals calculated for the process of differences between consecutive measurements of blood pressure displayed in Figure 1. The model of the process of differences D_i was estimated using first 20 observations of the monitored process of blood pressure measurements. Let us start our analysis from the case when residuals are calculated using the model estimated from the Yule-Walker equations. The estimated model is the AR model of the fourth order $AR(-0.9062, -0.6613, -0.1419, -0.0955)$ with the constant μ equal to 0.2953. The sample residuals calculated using this model have the mean value equal to 0.63, and standard deviation equal to 6.575. Hence, the upper control limit of the Shewhart control chart is equal to 20.356, and this chart is presented in Figure 4. The residuals displayed in Figure 4 are only slightly autocorrelated ($\rho(1) = 0.087$), so the estimated model well describes the considered model. However, sample residuals are moderately autocorrelated ($\rho(1) = 0.431$). Therefore, the parameters of the designed Shewhart control chart may be improved.

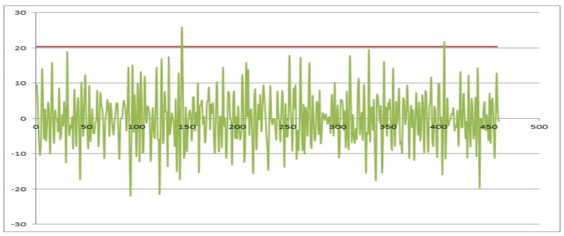


Figure 4. Control chart for first differences.

When the process model is estimated using Burg's algorithm the designed control chart is very similar. We have found that in this case the considered autoregression process is represented by the set of four parameters $(-0.987, -0.805, -0.217, -0.133)$. Then, residuals calculated for differences $D_5, \ldots, D_{19}$ have been used for the design of a control chart with the upper control limit equal to 20.29. The estimated model has been used for the calculation of residuals related to the next 80 observations. These residuals are displayed on the control chart. We can see that the monitored process seems to be under control, as all calculated residuals are located below the upper control limit.

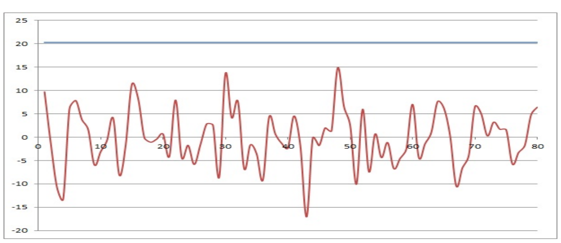


Figure 5. Control chart for residuals of differences of the first order related to measurements of blood pressure (Burg's method of prediction).

In comparison to a classical control chart for original observations, a control chart for residuals of differences has one important disadvantage: self-adaptation to a changed pattern of data. In order to explain this feature, let us transform our exemplary data by adding 20 to each observation starting from the 10th. The control chart in this case is presented in Figure 6.

From Figure 6, we can see that starting from the 10th point until the 12th point on the chart the values of displayed residuals sharply increase, but do not exceed the control limit. Later on, the process has returned to the previous level. It means that our chart is able to detect shifts of the monitored process only immediately after the jump. This is in sharp contrast to the classical Shewhart control chart (if it can be applied), where all data points observed after the shift indicate the deterioration of the monitored process. Thus, if the alarm is not generated immediately, it will be generated in the future quite randomly, despite the obvious deterioration of the monitored process. Therefore, we have to add an additional mechanism that will increase the probability of detection just after the shift.

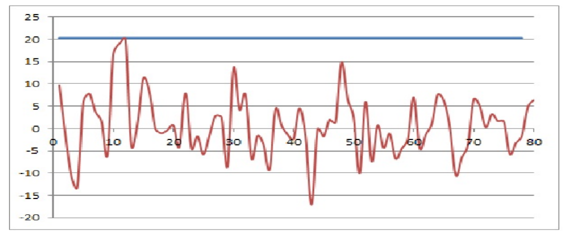


Figure 6. Control chart for residuals of differences of the first order related to measurements of blood pressure with a shifted process level,(Burg's method of prediction).

One of possible solutions of the problem mentioned above is to use an additional control chart. It can be a control chart for residuals calculated for second order differences defined as $D2_i = X_{i+2} - X_i$. The methodology for the design of this chart is exactly the same as that already described in this paper. Additional advantage of this approach is due to a fact that differences of the second order decrease or even cancel the impact of short cycles in the observed time series. A "weak" alarm signal is generated if it is generated on only one of these two charts. A "strong" alarm signal, that detects possible persistent deterioration, is generated when two consecutive points on the second chart are located beyond its control limits.

In our numerical example of shifted data, the model of the time series of differences of the second order, estimated from the sample of 20 observations, is the autoregression process of the second order $AR(-0.444, -0.555)$. Using this model, we can calculate residuals and design a respective control chart, presented in Figure 7. We can see that in the case of this control chart, deterioration of the process has been revealed with a delay of one measurement. Thus, if we have used both charts, we would detect the change of the process.

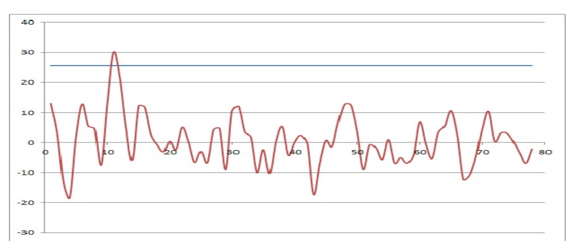


Figure 7. Control chart for residuals of differences of the second order related to measurements of blood pressure with a shifted process level (Burg's method of prediction).

Another possible solution, which is simpler for implementation, but theoretically less justified, is to calculate an additional residual as the difference between the observed difference of the second order and the predicted difference of the first order, but calculated for the previous observation, and to plot the maximum of these two residuals on the chart designed for the case of differences of the first order. A "weak" alarm is generated when a point on the chart is located beyond the control limits. For a "strong" alarm it is necessary to observe at least two consecutive points on the chart situated beyond the control limits.

It has to be stressed here, that the proposed procedures are based on rather heuristic reasoning, based on observations of a particular series of measurements. Unfortunately, closed mathematical formulae that describe statistical properties of a control chart when observed values of measurements are statistically dependent, as for now, do not exist (except for the simplest cases). Therefore, the properties of the proposed procedures have to be investigated in the future using complex simulation experiments.

IV. Using the sXWAM control chart for short process runs

One of the most important characteristics of a control chart is its rate of false alarms. An alarm is considered false if it is generated in a period of time when a monitored process is stable. False alarm rate is usually accompanied with good abilities to detect process disorders, so if this falsity does not lead to serious consequences, higher false alarm rates may be considered acceptable. However, when an alarm cannot be neglected because of its serious consequences, the false alarm rate should be very low. For example, in certain pharmaceutical production processes an alarm should trigger a stop of a process, and this may be very costly if the triggering alarm is false. In the case of a stable process, described by the model $AR(-0.987, -0.805, -0.217. -0.133)$ estimated from a sample of $n = 20$ observations, a chart presented in Figure 8 exhibits two false alarms.

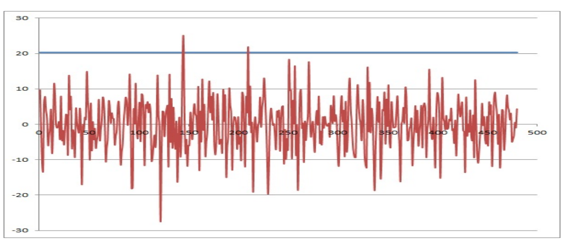


Figure 8. Control chart for residuals with two false alarms.

It has been observed by many authors (see [8], for more information) that control charts for autocorrelated data, especially those designed using small samples of observations, have elevated false alarm rates. Hryniewicz and Kaczmarek-Majer [8] have noted that this rather unfavorable property is somewhat related to the problem of bad predictability in short time series. Inspired by the very good properties of their prediction algorithm for short time series [20], they proposed in [8] a new control chart for residuals, named the XWAM control chart, based on the concept of model averaging.

Let us denote by M_0 the model of a monitored process estimated from a (usually) small sample, and describe its parameters by a vector $(a_{1,0}, \ldots . a_{p_0,0})$. We assign to this estimated model a certain weight $w_0 \in [0, 1]$. We also consider k alternative models $M_j, j = 1, \ldots, k$, each described by a vector of parameters $(a^0_{1,j}, \ldots, a^0_{p_j,j})$. In general, any model with known parameters can be used as an alternative one, but in this paper we restrict ourselves to the models chosen according to an extended version of the algorithm described in [13]. Let $w^{'}_1, \ldots, w^{'}_k$ denote the weights assigned to models $M_1, \ldots, M_k$ by this algorithm when only alternative models are considered. Because the total weight of the chosen alternative models is $1 - w_0$, to the estimated model we assign the weight w_0, and to each chosen alternative model we will assign a weight $w_j = (1 - w_0)w^{'}_j, j = 1, \ldots, k$.

When we model our process using $k+1$ models (one estimated from data, and k alternatives) each process observation generates $k+1$ residuals. In the case of differences of the first order considered in this paper, they are calculated using the following formula

$$\begin{aligned} z_{i,j} = \quad & d_i - (\mu + a_{1,j}d_{i-1} + \ldots + a_{p_j,j}d_{i-p_j}), \\ & j = 0, \ldots, k; i = p_j + 1, \ldots . \end{aligned} \tag{12}$$

In (12), we have assumed that for a model with $p_j, j = 0, \ldots, k$ parameters we need exactly p_j previous consecutive observations in order to calculate the first residual. Therefore, we need $i_{min} = \max(p_0, \ldots, p_k) + 1$ observations for the calculation of all residuals in the sample. For the calculation of the parameters of the XWAM control chart we use $n - i_{min} + 1$ weighted residuals calculated from the formula

$$z^{\star}_i = \sum_{j=0}^{k} w_j z_{i,j}, i = i_{min}, \ldots, n. \tag{13}$$

The central line of the chart is calculated as the mean of $z^{\star}_i$, and the control limits are equal to the mean plus/minus three standard deviations of $z^{\star}_i$, respectively. The operation of the XWAM control chart is a classical one. First decision is made after i_{min} observations. The weighted residual for the considered observation is calculated according to (13), and compared to the control limits. An alarm is generated when the weighted residual falls beyond the control limits.

The method for the construction of the XWAM chart was firstly proposed by Hryniewicz and Kaczmarek in [8] where they proposed an algorithm for the calculation of the weights of alternative models. This algorithm is based on the methods of computational intelligence, namely the DTW (Dynamic Time Warping) algorithm for comparison of time series. Unfortunately, this algorithm is computationally demanding, so in [13] they proposed its simplification, coined as the sXWAM (simplified XWAM). In this approach, Hryniewicz and Kaczmarek proposed not to compare original time series (observed and alternative), but their summarizations in terms of the autocorrelation functions of the pth order. Let $r_1, r_2, \ldots, r_p$ be the consecutive p values of the sample autocorrelation function, calculated using (5). Similarly, let $r_{1,i}, r_{2,i}, \ldots, r_{p,i}, i = 1, \ldots, J$ be the consecutive p values of the autocorrelation function of an alternative model. For given parameters of the alternative autoregression process $a_{1,i}, \ldots, a_{p,i}, i = 1, \ldots, J$ the values of $r_{1,i}, r_(2, i), \ldots, r_{p,i}, i = 1, \ldots, J$ can be found by solving the Yule - Walker equations (6). In general, the consecutive values of r_p can be computed using the following recursion equation

$$r_p = a_1 r_{p-1} + a_2 r_{p-2} + \cdots + a_p \tag{14}$$

Just like in [13], in this paper we consider only processes of the maximum fourth order. In such a case, explicit formulae for the first three autoregression coefficients are the following [13]:

$$r_1 = A_1, \tag{15}$$

$$r_2 = a_1 A_1 + a_2, \tag{16}$$

$$r_3 = \frac{a_1 B_1 + a_3 + (a_2 + a_4)(A_1 + A_2 B_1)}{1 - a_1 B_2 - (a_2 + a_4)(A_2 B_2 + A_3)}, \tag{17}$$

where

$$A_1 = \frac{a_1}{1 - a_2}, \tag{18}$$

$$A_2 = \frac{a_3}{1 - a_2}, \tag{19}$$

$$A_3 = \frac{a_4}{1 - a_2}, \tag{20}$$

$$B_1 = \frac{A_1(a_1 + a_3) + a_2}{1 - (a_1 + a_3)A_2 - a_4}, \tag{21}$$

$$B_2 = \frac{A_3(a_1 + a_3)}{1 - (a_1 + a_3)A_2 - a_4}. \tag{22}$$

Hence, the consecutive values of $r_4, r_5, \ldots$ can be directly computed from (14).

As the measure of distance between the estimated autocorrelations $r_1, r_2, \ldots, r_p$ and the correlations calculated for the ith alternative model $r_{1,i}, r_{2,i}, \ldots, r_{p,i}, i = 1, \ldots, J$ Hryniewicz and Kaczmarek-Majer [13] used a simple sum of absolute differences (called the Manhattan distance in the community of data mining)

$$dist_{i,MH} = \sum_{k=1}^{p} |r_k - r_{k,i}|, i = 1, \ldots, J. \tag{23}$$

In [1] Hryniewicz and Kaczmarek considered a slightly more general version of the sXWAM chart. As alternative models, they considered those autoregression models with k lowest values of $dist_{i,MH}$. Their weights, after some standardization, are inversely proportional to the distances of the closest models. The design of the sXWAM chart for residuals is thus much simpler than the original XWAM chart. The values of the autoregression functions for different alternative models can be computed in advance, and stored in an external file. This file can be read by a computer program, and used for choosing the model that fits to the observed sample (and its estimated autoregression function).

In this paper we try to improve the method used for finding alternative models. In order to do this we propose to use not only the autocorrelation functions ($ACF(p)$), but also the partial autocorrelation functions ($PACF(p)$). The kth value of the $PACF(p)$ function of the $AR(p)$ process, $\psi_p(k)$, is defined as the correlation between random variables X_t and X_{t+k} when the effect of the intermediate variables $X_{t+1} \ldots X_{t+k-1}$ that affect X_t and X_{t+h} has been removed. For the autoregression processes the $PACF(p)$ function has a very practical property. When the considered process is the AR process of the pth order, $AR(p)$, then all values of the $PACF(p)$ function are equal to zero for all $k > p$. The values of $\psi_p(k)$ can be found by solving the following system of equations

$$\begin{aligned} r_1 &= \psi_p(1) + \psi_p(2)r_1 + \ldots + \psi_p(p)r_{p-1} \\ r_2 &= \psi_p(1)r_1 + \psi_p(2) + \ldots + \psi_p(p)r_{p-2} \\ &\ldots \\ r_p &= \psi_p(1)r_{p-1} + \psi_p(2)r_{p-2} + \ldots + \psi_p(p) \end{aligned} \tag{24}$$

The system (24) is the so-called Toeplitz system of linear equations, and can be solved using the Durbin-Levinson recursion (see [14] for more information). The estimated values of the $PACF(p)$ functions can be calculated from (24) if we put estimated values of autocorrelations instead of their theoretical values. The distance between the estimated (from the sample) $PACF(p)$ function and the theoretical $PACF(p)$ of an alternative model is calculated using (23) with values of r_k and $r_{k,i}$ replaced, respectively, by ψ_k and $\psi_{k,i}$. Then, the sum of both distances, for the $ACF(p)$ and the $PACF(p)$ functions, is used for choosing the best alternative models and their weights.

V. Results and Discussion

In this section we present the results of computational experiments. In these experiments we use the original real-life data, and artificial data generated by adding some disturbances to the original data.

A. sXWAM control chart

The example of the sXWAM chart is presented in Figure 9 for the same original data that have been used for the construction of the control chart presented in Figure 8. For the design of this chart it was assumed that the weight for sample data is $w_0 = 0.7$. Five alternative process models have been found using the algorithm described above: $AR(-0.9, 0.5, 0.4, -0.3)$ with relative weight $w_1' = 0.201$, $AR(0.8, 0.7, -0.5, -0.3)$ with relative weight $w_2' = 0.201$, $AR(0.9, 0.5, -0.4, -0.3)$ with relative weight $w_3' = 0.200$, and $AR(-0.8, 0.7, 0.5, -0.3)$ with relative weight $w_4' = 0.199$, and $AR(0.8, 0.5, -0.3, 0.4)$ with relative weight $w_5' = 0.199$. We can see that in this case we have observed only one alarm generated at the same time point as one of the alarms generated on the control chart with non-weighted residuals. Experiments with artificially shifted process levels have shown that the detection ability of the proposed sXWAM chart are similar to that observed for the chart with non-weighted residuals.

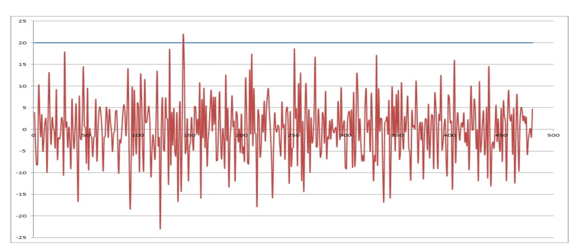


Figure 9. Control chart for weighted residuals with one false alarm.

B. Forecasting performance

Finally, we evaluate the forecasting performance of the proposed weighted model averaging approach with $ACF(p)$ and $PACF(p)$ (called $WAM*$, for short). We now present in detail the results obtained for the considered process of blood pressure measurements and the training database of AR processes of the first order. We assume that the model is estimated basing on a small sample ($n = 20$), and then, its verified for 6-step-ahead forecasts ($h = 1, 2, ..., 6$). The considered process has 480 observations that are used to create 454 samples. Each sample is divided into the first 20 observations that are used to estimate the predictive model, and the remaining 6 observations are used to verify its predictive performance. Alternative models and forecasts are calculated for each sample.

For example, the first sample is created for the first 26 observations of the process. The $WAM*$ approach delivers the following best alternative models and their weights: $AR(-0.4), w_1 = 0.34$, $AR(-0.5), w_2 = 0.33$, $AR(-0.3), w_3 = 0.33$. For the next sample (starting with n=2 element of the blood pressure measurements and ending with n=21 element of this process), exactly the same best alternative models are calculated. Slightly different models are calculated for the fourth sample (starting with n=4 element of the blood pressure measurements and ending with n=23 element of this process): $AR(-0.3), w_1 = 0.35$, $AR(-0.2), w_2 = 0.34$, $AR(-0.1), w_3 = 0.31$. In Table I, the frequency of the applied alternative models is presented.

TABLE I. Frequency of alternative models calculated with the $WAM*$ approach within all samples of the process with blood pressure measurements.

	M_1	M_2	M_3
AR(0.2)	-	0.00	0.00
AR(0.1)	0.00	0.00	0.01
AR(0.0)	0.01	0.02	0.05
AR(-0.1)	0.06	0.04	0.07
AR(-0.2)	0.07	0.14	**0.21**
AR(-0.3)	0.18	**0.27**	**0.24**
AR(-0.4)	**0.39**	**0.21**	0.15
AR(-0.5)	**0.19**	0.19	0.14
AR(-0.6)	0.07	0.06	0.11
AR(-0.7)	0.03	0.03	0.02
AR(-0.8)	0.01	0.03	0.00

As observed in Table I, the most frequent best alternative model is $AR(-0.4)$, selected in 39% of cases. As the second best model, $AR(-0.3)$ is selected in 27% of the cases. As the third best alternative model, again $AR(-0.3)$ is selected in 24% of the cases.

Finally, we summarize the forecasting results to evaluate the accuracy of the proposed approach. The following measures are applied for evaluation: the mean absolute error (MAE), the standard deviation of the MAE (stdD-MAE) the mean squared error (MSE) and the standard deviation of the MSE (stdD-MSE). As observed, we compute the commonly used scale-dependent measures (MAE, MSE, MDAE) because they are useful when comparing different methods applied to the same set of data, and this is our case.

Table II shows the accuracy measures depending on the forecast horizon. It is observed that in all cases, the proposed $WAM*$ method delivers forecasts that are more accurate than the forecasts calculated according to the AR process estimated with the Yule-Walker equations.

TABLE II. Forecast accuracy (the horizon up to 6-step-ahead) measured with MSE, stdD-MSE, MAE and stdD-MAE according to the proposed $WAM*$ approach and the respective estimated AR process.

Fcst. hor.	Method	MSE	stdD-MSE	MAE	stdD-MAE
1	$WAM*$	57.32	85.92	5.92	4.72
1	est AR	60.08	92.59	6.07	4.83
2	$WAM*$	64.82	76.00	6.30	4.05
2	est AR	68.91	81.64	6.51	4.12
3	$WAM*$	66.51	65.69	6.39	3.46
3	est AR	70.01	69.56	6.58	3.52
4	$WAM*$	67.21	58.56	6.43	3.06
4	est AR	69.96	61.46	6.57	3.12
5	$WAM*$	67.49	52.76	6.44	2.75
5	est AR	69.77	54.89	6.56	2.79
6	$WAM*$	**67.64**	**48.21**	**6.45**	**2.52**
6	est AR	**69.64**	**49.79**	**6.56**	**2.55**

For example, the average MSE of the 6 forecasts for 6-steps-ahead amounts to 67.64 according to the proposed approach and to 69.64 for the estimated process. For the comparative purposes, we calculate the errors obtained for the proposed model averaging approach in comparison to the errors of the forecasts calculated from the process estimated basing on the Yule-Walker equations (referenced as 'est AR' forecast). The relative errors for the aforementioned measures are presented in Table III.

TABLE III. Relative change of MSE, stdD-MSE, MAE and stdD-MAE for forecasts (up to 6-step horizon) of the proposed $WAM*$ approach compared to the respective estimated AR process.

Fcst. hor.	Method	R-MSE	R-stdD-MSE	R-MAE	R-stdD-MAE
1	$WAM*$	0.95	0.93	0.98	0.98
2	$WAM*$	0.94	0.93	0.97	0.98
3	$WAM*$	0.95	0.94	0.97	0.98
4	$WAM*$	0.96	0.95	0.98	0.98
5	$WAM*$	0.97	0.96	0.98	0.98
6	$WAM*$	**0.97**	**0.97**	**0.98**	**0.99**

As observed in Table III, the proposed model averaging $WAM*$ approach enables to improve all forecasts estimated with the traditional Yule-Walker equations. The improvement ranges from 0.93 to 0.99 depending on the considered accuracy measure.

C. Control chart for residuals of differences of the first order

Now, let us show how this new improved method of prediction can be utilized in the construction of a control chart. For calculation of residuals of first differences we will use predictions calculated according the autoregression model (4). The estimated process model minimizes the modified Akaike's BIC criterion, calculated according to (10). Suppose, that our control chart has to be put into operation after observing first 20 measurements of blood pressure. Hence, our sample of first differences consists of 19 observations. The optimal (with respect to the BIC criterion) model, estimated from this sample, is the following: $\mu_0 = 0.1642$ and $a_{0,1} = -0.559$. Thus, our residuals a described by the autoregression model of the first order. Let us assume now that our alternative models have to be chosen from among autoregression models of first and second order. In this case, three best alternative models are the following: ($\mu_1 = 0.263, a_{1,1} = -0.9, a_{1,2} = -0.6$) with the relative weight 0.371, ($\mu_2 = 0.232, a_{2,1} = -0.8, a_{2,2} = -0.5$) with the relative weight 0.359, and ($\mu_3 = 0.253, a_{3,1} = -0.9, a_{3,2} = -0.5$) with the relative weight 0.269. Because all our alternative models are of the second order, we can calculate only 18 prediction of first differences, and thus only 18 residuals. When we take into account only the estimated model, the mean value of calculated residuals is equal to 0.673, and sample standard deviation is equal to 10.393. Hence, the Shewhart two-sided control chart for the mean value, calculated using only estimated model, has the following control limits: $LCL = -30.50$, and $UCL = 31.85$. When we use the estimated model for the calculation of residuals for the remaining 460 measurements, the control chart will look like this presented in Figure 10. As we can see, both control lines are rather distant from the charted process values. Thus, this chart may not be effective in the detection of process instability.

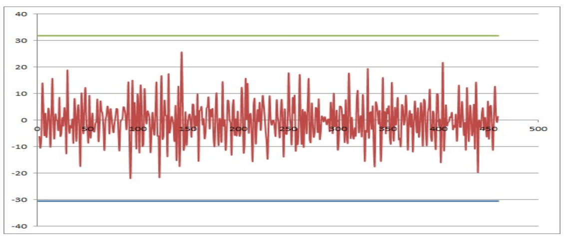


Figure 10. Control chart for residuals of differences of the first order related to measurements of blood pressure (estimated AR model used for prediction).

Now, let us consider a chart designed using only weighted residuals calculated according to the chosen alternative models. In this case, the mean value of calculated residuals is equal to 0.303, and sample standard deviation is equal to $6,628$. Hence, the Shewhart two-sided control chart for the mean value, calculated using only estimated model, has the following control limits: $LCL = -19.58$, and $UCL = 20.19$. Note, that in this case the standard deviation is significantly smaller, and the control lines a closer to the process observations. When we use the weighted alternative models for the calculation of residuals for the remaining 460 measurements, the control chart will look like this presented in Figure 11. We can see from Figure 11 that in this case four points on the chart fall beyond the control limits. Thus, we observe four alarms, and two of them (these showing significant increase of blood pressure) are definitely false. Moreover, insignificant (from a medical point of view) shifts of the monitored process may cause many false alarms.

Finally, let us use the $WAM*$ approach, and construct a chart for which we assign the weight w_0 to the estimated model, and the weight $1 - w_0$ to the alternative models. Suppose, that we take $w_0 = 0.5$. In this case, the mean value of calculated residuals is equal to 0.3488, and sample standard deviation is equal to $8,294$. Hence, the Shewhart two-sided control chart for the mean value, calculated using the $WAM*$ approach, has the following control limits: $LCL = -24.39$, and $UCL = 25.37$. When we use the $WAM*$ approach for the calculation of residuals for the remaining 460 measurements, the control chart will look like this presented in Figure 12.

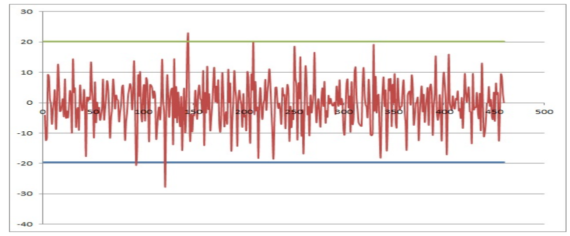


Figure 11. Control chart for residuals of differences of the first order related to measurements of blood pressure (weighted alternative AR models used for prediction).

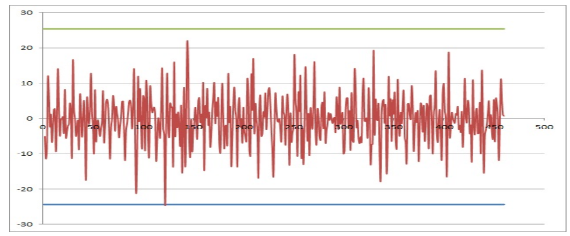


Figure 12. Control chart for residuals of differences of the first order related to measurements of blood pressure ($WAM*$ approach used for prediction).

The control chart displayed in Figure 12 represents a compromise between a chart designed using only an estimated model, and a chart designed using alternative models. On the one hand, in contrast to the chart presented in Figure 10, it should react to dangerous shifts of blood pressure, as the control limits are not so wide, as in the case presented in Figure 10. However, on the other hand, in contrast to the chart presented in Figure 11, where the control limits seem to be too narrow, no alarm signals are displayed for this seemingly stable process. It means that the control limits are not too narrow, and the chart should not generate too many false alarms in the case of insignificant variations of the monitored process. One has to stress here, that the behavior of a control chart presented in Figure 10 may lead to dangerous consequences (no necessary alarm signals). Observing too many false alarms is not so dangerous, but may decrease reliance on the used control chart. Therefore, in statistical process control we are always looking for compromise solutions, such as that presented in Figure 12.

In the design of our control chart we have used AR models of maximal second order. One could ask, however, about a similar design using models of higher order. In our calculations we also considered models of the fourth order. Predictions using such models are definitely more accurate. However, when the number of available measurements is, as in our case, strongly limited, then the effective sample size (the number of calculated residuals) becomes smaller. For example, when we use models of the fourth order, the sample size decreases to 15. Thus, the total variability of residuals may even increase, and the designed chart may be less effective.

VI. Conclusions

This paper has to be considered as a methodological one. As an illustrative example we have considered monitoring stability of a short and non-stationary processes using a simple tool such as a Shewhart control chart. We consider a monitored process stable, if its predicted future values are not very different from the observed ones. Such processes are often encountered in practice, and in this paper are represented by a certain health-recovery process, where natural randomness of measured health-related characteristics is accompanied by random or deterministic trends. Statistical analysis of non-stationary processes is usually very difficult and costly for implementation, as it requires large amount of available data and sophisticated specialized software. It can be used at intensive-care hospital units or in cases when patient's life is endangered. However, in many cases, it is completely sufficient to monitor the state of health using personal measuring devices, and to alarm a patient (or his/hers physician) only in the case of unexpected events. We are of the opinion that this can be done using simple tools, like Shewhart control charts. Such procedures using simple software implemented, e.g., in personal measurement equipments. The chart proposed in this paper should definitely fulfill such requirements of simplicity.

In our research, we assume, as it is often done in statistical process control (SPC), that at its initial stage the monitored process is supervised (e.g., by a physician), and considered as stable. Data from this stable period, considered as our sample data, are used for the identification of the monitored process and construction of a control chart. Because simple methods for monitoring non-stationary processes do not exist, we propose to monitor differences of the first order (i.e,, differences between values of consecutive measurements). This approach is effective for linear or approximately linear trends. When we consider, as in this paper, short series of observations, this assumption seems to be rather realistic. However, it is possible to apply the proposed methodology for differences of a higher order. For example, in this paper, we also consider differences of the second order which can be used in the case of processes with alternating (e.g., morning and evening) process levels. In our investigations we have assumed that our series of observations are rather short, and the monitored process has to be identified using a small sample of measurements. This assumption reflects reality when health-recovery processes is evaluated by a physician for only short time, and the period in which the process has to be stable is also short (e.g., until a next treatment is applied). For this reason, we have proposed a novel statistical tool, sXWAM chart, developed recently by us, and the new chart, designed using the $WAM*$ approach, proposed by us in this paper.

The performance of the proposed method has been verified using real-life data obtained from a patient recovering from a hypertension episode, who measured his blood tension once a day for a period of 480 consecutive days. Statistical analysis of these data has shown that basic assumptions about statistical independence of consecutive measurements, which are used in the design of control charts applied in the monitoring of hypertension patients, and used in the procedures described in Solodky et al. [21], and Hebert and Neuhauser [22], have not been fulfilled. Despite the limited amount of data considered in this research, the presented results show that the proposed method is promising. The proposed methodology is general, and can be used for the analysis of any type of health-related measurements. One has to admit, however, that the particular prediction model - described by the equation (2) - is too simple for a proper description of continuous 24/7 measurements, either in clinical environment or taken from wearable sensors. In such applications much more complicated models, that take into account, e.g., periodical changes of a measured process level, must be used. A good presentation of this problem can be found in the analysis of the results of the 10th Annual PhysioNet/Computers in Cardiology Challenge 2009, which was devoted to predicting the Acute Hypotension Episodes (AHE), and described in the paper by Moody and Lehman [23]. Its participants provided various sophisticated solutions such as: neural networks, a rule-based approach, decision trees or support vector machines. A short review of other recent approaches for the AHE prediction can be also found in the paper by Jiang et al. [24]. Some of these methods may be used in our approach as replacements for our prediction model (2). Unfortunately, their computational complexity is currently too high for such personal devices like tablets or smartphones. Therefore, their usage in our approach is possible, but only in the case of external data processing by sufficiently powerful computers.

In our current research (with participation of clinical psychiatrists) we apply the proposed methodology for the analysis of self-assessment data provided by patients suffering from psychiatric bipolar disorder. It is known from the investigations described recently by Bonsall et al. [25], Maxhuni et al. [26], and Vasquez-Montes et al. [27] that classical control charts cannot be used for monitoring the behavior of such patients. Preliminary results of the application of the procedure described in this paper, described in Kaczmarek-Majer et al. [28], are considered by psychiatrists as very promising, and show great potential of the proposed methodology.

Finally, one has to note, that despite great progress in the application of statistical process control (SPC) in health care the proposed solutions, like the one presented in this paper, are still in an experimental stage. Therefore, it will take some time before their application in medical devices will be approved by regulatory bodies.

REFERENCES

[1] O. Hryniewicz and K. Kaczmarek-Majer, "Monitoring of health-recovery processes with control charts," in Proceedings of ACCSE 2017 : The Second International Conference on Advances in Computation, Communications and Services, P. Lorenz and J. Trienekens, Eds., 2017, pp. 6 – 11.

[2] D. Montgomery, Introduction To Statistical Quality Control (6th Edition). J.Wiley, 2011.

[3] W. Woodall, "The use of control charts in health-care and public-health surveillance," Journal of Quality Technology, vol. 38, no. 2, 2006, pp. 89 – 104.

[4] R. Tennant, M. Mohammed, J. Coleman, and U. Martin, "Monitoring patients using control charts: a systematic review," International Journal for Quality in Health Care, vol. 19, 2007, pp. 187–194.

[5] J. Thor, J. Lundberg, J. Ask, J. Olsson, C. Carli, K. Härenstam, and M. Brommels, "Application of statistical process control in healthcare improvement: systematic review," Quality Safety in Health Care, vol. 16, 2007, pp. 387–399.

[6] P. Winkel and N. Zhang, "Statistical process control in clinical medicine," in Statistical Methods in Healthcare, F. Faltin, R. Kenett, and F. Ruggeri, Eds. J. Wiley, 2012, ch. 15.

[7] M. Møller and J. Anhøj, "Statistical process control in healthcare improvement new kid on the block?" Acta Anaesthesiologica Scandinavica, 2018.

[8] O. Hryniewicz and K. Kaczmarek-Majer, "Monitoring of short series of dependent observations using a XWAM control chart," in Frontiers in Statistical Quality Control 12, S. Knoth and W. Schmid, Eds. Springer, 2017, p. (in press).

[9] G. Cornélissen, F. Halberg, D. Hawkins, K. Otsuka, and W. Henke, “Individual assessment of antihypertensive response by self-starting cumulative sums,” Journal of Medical Engineering & Technology, vol. 21, 1997, pp. 111–120.

[10] T. Albloushi, A. Suwaidi, N. Zarouni, A. Abdelrahman, and M. Shamsuzzaman, “Design of $\bar{X}\&R$ control charts for monitoring of care for hypertension,” in Proc. of IEOM 15, 2018.

[11] Y. Jung and H. Kim, “Detection of pvc by using a wavelet-based statistical ecg monitoring procedure,” Biomedical Signal Processing and Control, vol. 36, 2017, pp. 176–182.

[12] T. Lambeth, M. Rojas, A. Holmes, and R. Dail, “First golden hour of life: A quality improvement initiative,” Advanced Neonatal Care, vol. 16, 2016, pp. 264–272.

[13] O. Hryniewicz and K. Kaczmarek-Majer, “Monitoring series of dependent observations using the sxwam control chart for residuals,” in Soft modelling in industry, ser. Studies in Systems, Decision and Control, P. Grzegorzewski and A. Kochanski, Eds. Springer, 2017 (in press).

[14] P. Brockwell and R. Davis, Introduction to Time Series and Forecasting, 2nd ed. New York: Springer, 2002.

[15] D. Kwiatkowski, P. Philips, P. Schmidt, and Y. Shin, “Testing the null hypothesis of stationarity against the alternative of a unit root,” Journal of Econometrics, vol. 54, 1992, pp. 159–178.

[16] G. Box, G. Jenkins, and G. Reinsel, Time Series Analysis. Forecasting and Control. Hoboken NJ: J.Wiley, 2008.

[17] C. Collomb, “Burg’s method, algorithm and recurssion,” Tech. Rep., 2009, retrieved: April, 2017.

[18] H. Akaike, “Time series analysis and control through parametric model,” in Applied Time Series Analysis, D. Findley, Ed. New York: Academic Press, 1978, pp. 1 – 23.

[19] L. Alwan and H. Roberts, “Time-series modeling for statistical process control,” Journal of Business & Economic Statistics, vol. 6, 1988, pp. 87 – 95.

[20] O. Hryniewicz and K. Kaczmarek-Majer, “Bayesian analysis of time series using granular computing approach,” Applied Soft Computing Journal, vol. 47, 2016, pp. 644–652.

[21] C. Solodky, H. Chen, P. Jones, W. Katcher, and D. Neuhauser, “Patients as partners in clinical research: a proposal for applying quality improvement methods in patient care,” Medical Care, vol. 36 (Suppl.), 1998, pp. AS13–20.

[22] C. Hebert and D. Neuhauser, “Improving hypertension care with patient-generated run charts: physician, patient, and management perspectives,” International Journal of Quality Management in Health Care, vol. 13, 2004, pp. 174–177.

[23] G. Moody and L. Lehman, “Predicting acute hypotensive episodes: The 10th annual physionet/computers in cardiology challenge,” Computers in Cardiology, vol. 36, 2009, pp. 541–544.

[24] D. Jiang, L. Li, B. Hu, and Z. Fan, “An approach for prediction of acute hypotensive episodes via the hilbert-huang transform and multiple genetic programming classifier,” International Journal of Distributed Sensor Networks, 2015.

[25] M. Bonsall, S. Wallace-Hadrill, J. Geddes, G. Goodwin, and E. Holmes, “Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder,” Proceedings of the Royal Society B, vol. 279, 2012, pp. 916–924.

[26] A. Maxhuni, A. Muñoz-Meléndez, V. Osmani, H. Perez, O. Mayora, and E. Morales, “Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients,” Pervasive and Mobile Computing, vol. 31, 2016, pp. 50–65.

[27] M. Vazquez-Montes, R. Stevens, R. Perera, K. Saunders, and J. Geddes, “Control charts for monitoring mood stability as a predictor of severe episodes in patients with bipolar disorder,” International Journal of Bipolar Disorders, vol. 6:7, 2018.

[28] K. Kaczmarek-Majer, O. Hryniewicz, K. Opara, W. Radziszewska, A. Olwert, J. Owsiński, and S. Zadrożny, “Control charts designed using model averaging approach for phase change detection in bipolar disorder,” in Uncertainty Modelling in Data Science, S. Destercke, T. Denœux, M. Gil, P. Grzegorzewski, and O. Hryniewicz, Eds. Springer International, 2018 (in press).

Modeling of Driver's Distraction State based on Body Information Analysis

Kazuhito Sato, Masafumi Sawataishi, Hirokazu Madokoro
Department of Machine Intelligence and Systems Engineering,
Faculty of Systems Science and Technology, Akita Prefectural University
Yurihonjo, Japan
ksato, m18a016, madokoro @akita-pu.ac.jp

Momoyo Ito
Institute of Technology and Science,
Tokushima University
Tokushima, Japan
momoito@is.tokushima-u.ac.jp

Sakura Kadowaki
Smart Design Corp.
Akita, Japan
sakura@smart-d.jp

***Abstract* - For this study, we defined a "concentration state" as that when a driver performs only driving tasks and a "distraction state" as that when a driver performs a driving task and a mental arithmetic task simultaneously. Based on results of these driving tests, we elucidate the characteristics of safety confirmation behaviors by near-misses according to differences between two driving conditions when approaching an intersection. Specifically, using Bayesian Networks (BNs) to express the relation between safety confirmation behaviors graphically for driving scene and driver's internal state, we analyze correlations between characteristic body information (i.e., eye-gaze / face orientation) and operation information (i.e., steering wheel, accelerator, and brake) when switching from a "concentration state" to a "distraction state" based on viewpoints of driving style and driving workload sensitivity. Using evaluation experiments, by constructing internal state estimation models of subjects for whom driving style and driving workload sensitivity are mutually opposite, we try to analyze body information and the operation information by which the influence of a distracted state appears easily through the specification of unique behavior patterns associated with each driving scene.**

Keywords - body information; driver behavior; human engineering; near miss; non-regulated intersection

I. INTRODUCTION

According to accident statistics, at non-regulated intersections where numerous crossing accidents occur, non-confirmation of safety is common: cognitive distraction states account for approximately 3/4 of human error in crossing accidents. Of pedestrian accident factors that result in death, 35% are states of careless driving, classified as a "distraction state" or "state of fatigue" according to the level of driver arousal. To clarify the relevance between a "distraction state" and a driver's driving characteristics, we have been studying a possibility of modeling drivers' internal states when switching from a concentrating state to a distracted state [1]. Detection indices of the "distraction state" are restricted to physiological information and body information. Methods of detecting attentiveness and arousal level from face orientations and eyelids (e.g., blinking or eyelid opening) captured by an onboard camera [2] [3] and an approach to detect the reduction of driver arousal level from vehicle behaviors [4] have been put to practical use as techniques of fatigue state estimation for drivers. More recently, research and development projects to detect and use eye-gaze movements have been promoted actively for the detection of signals of inattentive driving and distraction states [5] [6]. Furthermore, as case studies dealing with facial expression changes during driving, studies dealing with the relation between the evaluation value of facial expression changes by a third party and the degree of fatigue have been reported as attempts at driver fatigue measurements [7] [8]. In these studies, various methods been attempted, with increasing capabilities demonstrated by the evaluation of facial expression changes.

In earlier studies [9] [10], to extract characteristic behavior patterns that occur when a driver becomes distracted, we used a mental arithmetic task to perform a driving experiment simulating a driver's distraction state. Then, we analyzed the relation between data of physical information such as the face orientation / eye-gaze movement obtained from experiments and a driver's driving characteristics. Results show that a clear difference might occur in the distribution of eye-gaze movement in a concentration state and distraction state. Furthermore, through analysis of correlation between the physical information and operating information of vehicles, we confirmed the possibility of modeling drivers' internal states when switching from a concentrating state to a distracted state.

For this study, we attempt to model driving behavior using Bayesian Networks (BNs) [11] for a dataset acquired in earlier research. Actually, BNs can represent a causal relation of several random variables as a graph structure. Furthermore, one can obtain the *a posteriori* probability distribution of other variables by setting the evidence of the causal variable. For this study, we set the evidence for the driving scene and the driver's internal state, and perform probabilistic inference. Then, we find a posterior probability distribution of each node of physical information, biometric information, and operation information, and analyze the characteristic driving behavior, which changes depending on the driving scene and the driver's internal state.

This paper is presented as follows. We review related work to clarify the position of this study in Section II. Section III presents a definition of the running route and near-miss events assuming the sudden appearance of a bicycle. In Section IV, we examine two cases of "concentration state" to assess driving tasks and "distraction state" to assess driving tasks and mental arithmetic tasks simultaneously. Additionally, we identify characteristic behavior patterns corresponding to the three driving scenes and analyze the correlation between physical information and operation information, each of which is likely to be affected by the distracted state using the constructed internal state estimation model. Finally, we present conclusions and intentions for future work in Section V.

II. Related Works

A cognitive distraction state is difficult to ascertain from external appearances because it is an internal state of a driver. No estimation method has been established. Proposed methods of distraction state detection [12] were classifiable into four types based on the modalities to be measured [13] [14]. Considering the burden on drivers, the use of "physiological information" and "subjective evaluation" of drivers is unrealistic for application to actual driving conditions. For "operating information" such as steering and braking, which are involved directly with risk, notifying the driver even after detecting the distraction state can occur too late for danger avoidance. The distraction states addressed in this study include a "thinking state" and a "blank state" in which attention resources are distributed, and "impatience, or a frustrated state" attributable to time constraints under driving tasks. Abe et al. examined a "thinking state" [15]. Honma et al. reported a "blank state" [16]. Each study examined the experimental verification of each state. Each confirmed the presence of distinctive situations in which oversight and delayed discovery of changes in the surrounding circumstances would occur easily.

In earlier studies of modeling driving behaviors, Kumano et al. modeled eye-gaze movement using Dynamic Bayesian Networks (DBNs) [17]. They constructed a model of eye-gaze behaviors using DBNs from elements such as body movements, driving operations, and driving situations for identifying three types of eye-gaze movements such as forward gazing, dead angle confirmation, and inattention. As a result of verifying observation parameters because of differences in eye-gaze behaviors, they demonstrated the possibility of acquiring a high discrimination rate using context information such as driving operations and driving situations. However, the internal state estimation of each driver has not been reached for prediction of risky driving.

However, in constructing real-time distraction detection systems, Ahlstrom et al. investigated the usefulness of a real-time distraction detection algorithm called AttenD [18]. They described that it is difficult to measure the driver distraction state. There is no commonly accepted ground truth for driver distraction [19]. The validation of AttenD was performed in the field using the general methodological setup of a field operational test. That earlier study found that true distraction is difficult to attain in an artificial setting such as a simulator. However, AttenD seems not to respond appropriately our target distraction states such as a "thinking state" or "blank state" as described above. Therefore, we are convinced of the necessity of modeling along with each cycle of recognition, judgment, and operation which is the basis of driving tasks.

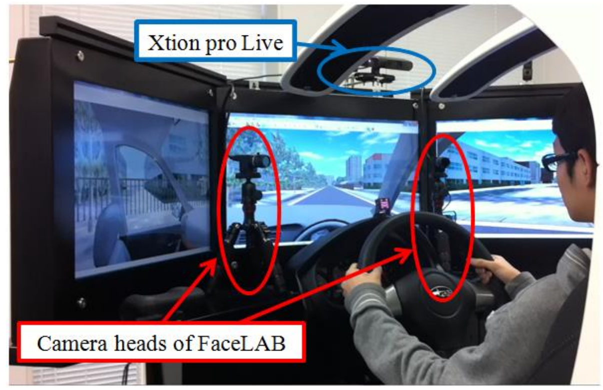


Figure 1. Experimental system for measuring driver behaviors.

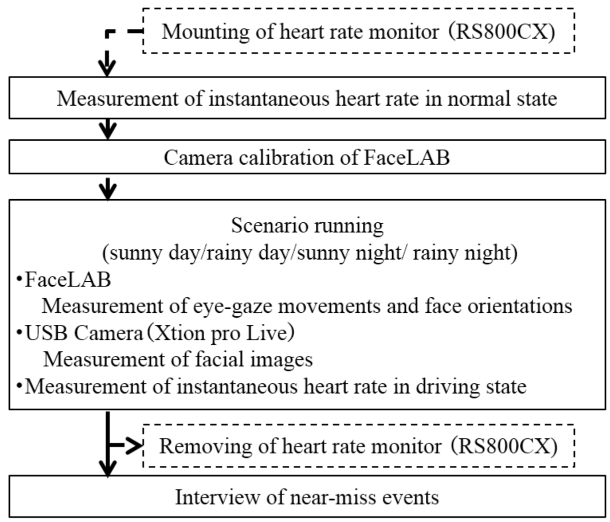
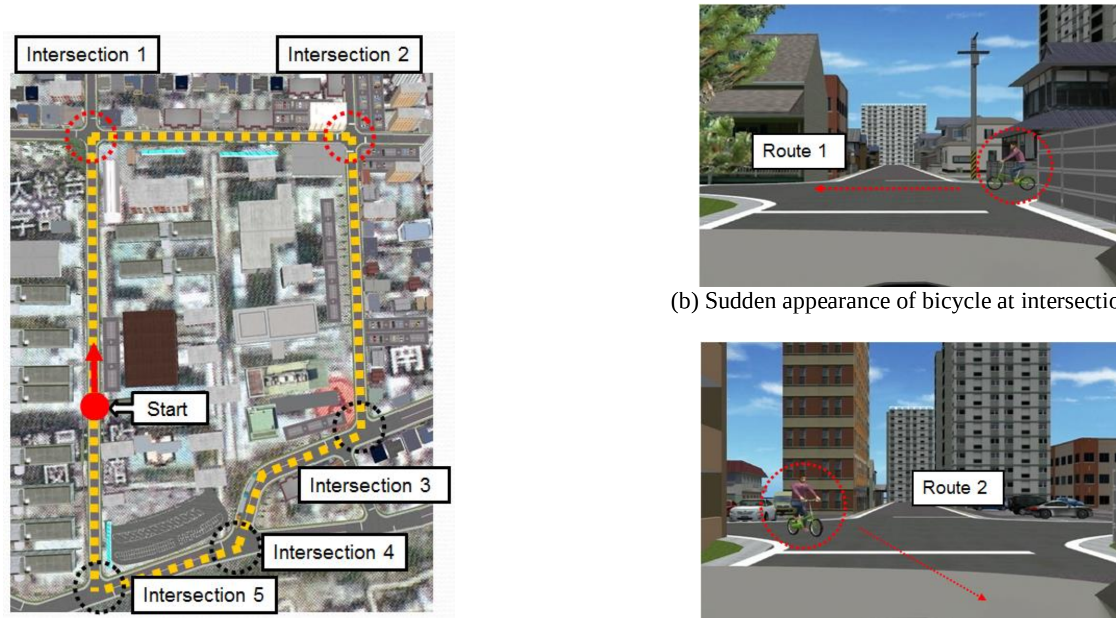


Figure 2. Outline of experimental protocols.



(a) Running route for simulation

(b) Sudden appearance of bicycle at intersection 1

(c) Sudden appearance of bicycle at intersection 2

Figure 3. Simulation course with near-miss events of two types (Route 1/Route 2).

III. METHODS

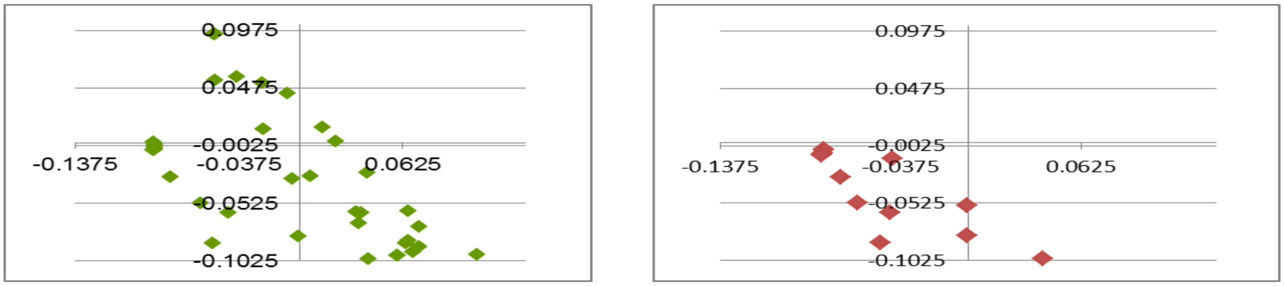
A. Experimental Systems

This study used a driving simulator (DS) to assess driving behaviors for freely set road environments and traffic conditions that affect driver behaviors. Figure 1 portrays the experimental system configuration used to measure driver behaviors. The DS used for experiments has platforms corresponding to compact and six-axis motion, which is equipped with ordinary cars. The DS has three color liquid crystal displays mounted in the front of the cabin, and has a function reproducing pseudo-driving environments that are freely configurable to horizontal viewing angles. As Figure 1 shows, to measure body information such as head poses, face orientations, and eye-gaze movements without restraining drivers, we installed cameras to the left and right in the center of the three-color liquid crystal monitors mounted in front of the cabin. Additionally, we set an infrared pod on top of the instruments in front of the cabin. Here the camera heads and infrared pod are input-based sensors of a head-gaze tracking device (FaceLAB; Ekstreme Machine Corp.).

B. Experimental Protocols

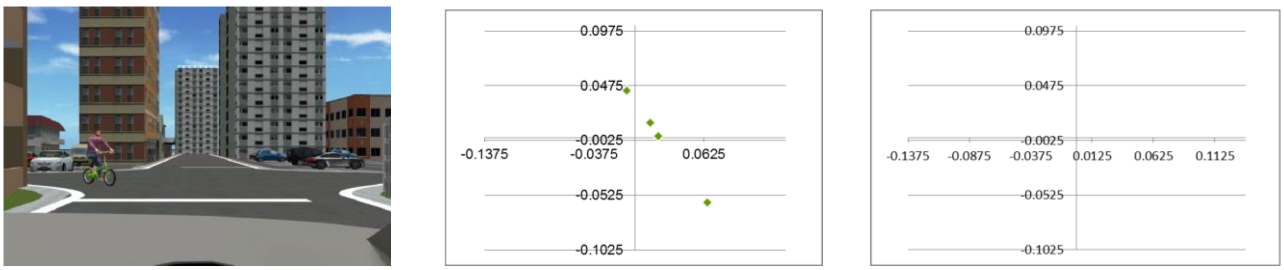
Figure 2 depicts the experimental protocol outline. The "concentration state" for driving represents driving states while performing only driving tasks. In contrast, the "distraction state" is defined as a "thinking state" during which driving tasks and simple mental arithmetic tasks are performed simultaneously. Similarly to experiments described by Abe et al. [15], mental arithmetic tasks used in our experiment included one digit addition, presented to drivers at 3 s intervals. Mental arithmetic results were to be reported verbally. The correctness of the results of each mental calculation was not reported to the driver during driving. Initially, as individual characteristics of each subject, we conducted an examination of the following questionnaire methods: attitude, oriented, and concept to work on driving were performed using the driving style check sheet [20]. Regarding the types of operation burdens that were strongly felt, they were performed using a driving load sensitivity check sheet [21]. In one running test, for each target subject wearing a heart rate monitor (RS800CX; Polar), we measured the instantaneous heart rate of a normal state during 1 min in advance. Next, to improve the measurement accuracy for face orientations and eye-gaze movements of each participant, we calibrated the cameras of the head-gaze tracking device (FaceLAB). We recorded a face video while driving with the USB camera (Xtion Pro Live; ASUS Corp.) to analyze the facial expressions of the subject. After these preparations, each participant ran three laps along the running scenario described later in Section Ⅲ.C, by synchronizing the time base of each measuring device. Finally, using a questionnaire that specifically examines traffic events occurring at intersections, subjective reviews, a four-stage check, were also conducted when a near-miss occurred.

After obtaining the approval of Akita Prefectural University Research Ethics Board, the experiment contents for all subjects were explained fully to participants in advance. We obtained written consent of participants. From



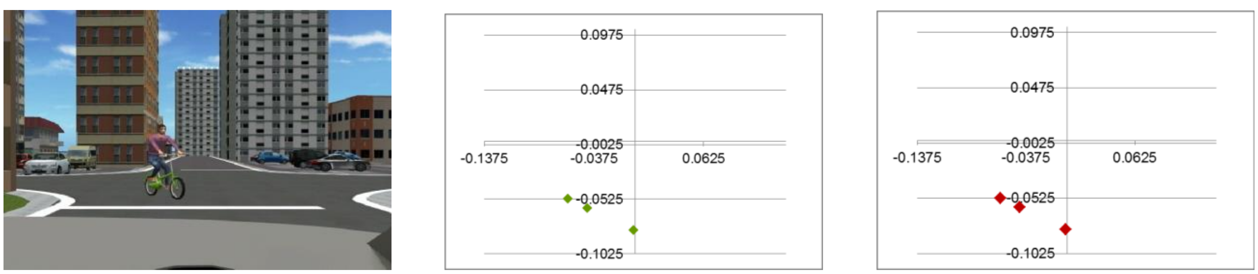
(a) Eye-gaze movements (b) Face orientations

Figure 4. Scatter diagram of the sections ranging from detection to gaze-tracking of sudden appearance of bicycle.



(a) Target scene (b) Eye-gaze movements (c) Face orientations

Figure 5. Detection case of sudden appearance of the bicycle for the first time.



(a) Target scene (b) Eye-gaze movements (c) Face orientations

Figure 6. Case in which eye movements and face orientations were plotted concurrently.

each, we also obtained an agreement to publish a face image along with the consent to experimental participation.

C. Near-miss events and running scenarios

Figure 3 presents a definition of the running route and near-miss events assuming the sudden appearance of a bicycle. The running route is a circuit course simulating non-controlled intersections near the Tokushima University campus, where joint research has been conducted. As targets, we set two types of sudden appearance of bicycles for intersection 1 and intersection 2 as shown in Figure 3. One near miss is defined as the sudden appearance of a bicycle along route 1, which is crossing from the right side to the left side of the vehicle. The other near-miss entails the sudden appearance of a bicycle along route 2, which is interfering with the vehicle path at intersection 2. With the main campus gate as a starting point, each subject traveled from the intersection 1 in order through intersection 5. In addition, intersection 1 and intersection 2 are non-controlled intersections that had poor visibility.

Subsequently, we present an overview of running scenarios in the following. The basic traveling scenario is configured to three laps of the running route described above. We conducted control of near-miss events for generation as follows. The first lap and the third lap were without near-miss events. The second lap includes the sudden appearance of bicycles along route 1 at intersection 1 and route 2 at intersection 2. Furthermore, it is necessary to control other traffic flows such as traffic turning right and straight oncoming vehicles at intersection 1, and that of cars crossing in front of one's own vehicle at intersections 2 and

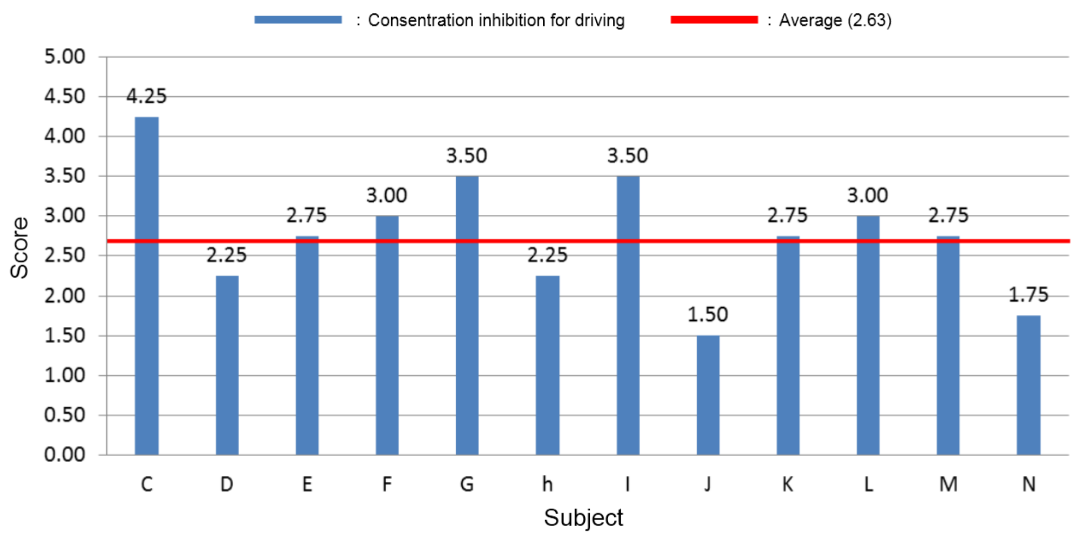


Figure 7. Calculation results of "concentration inhibition for driving" for all participants.

3. Based on the fundamental running scenarios described previously, we must assess behaviors in different weather (i.e., sunny or rainy) and time periods (i.e., daytime or nighttime). We prepared four driving environments: sunny day, rainy day sunny night, and rainy night.

IV. Experiment

For experiments, we examined two cases of "concentration state" to assess driving tasks and "distraction state" to assess driving tasks and mental arithmetic tasks simultaneously. Based on the running scenarios described in Section III.C, the experiments were conducted to three laps of the same route with one running environment, for driving environments of four types with varying weather and time zones. Furthermore, by changing the driving environments in the order of sunny day / rainy day / sunny night / rainy night, we examined the same conditions for all participants, all of whom were students of our university with ordinary vehicle driver licenses. These 10 men were designated as C, D, E, H, I, J, K, L, M, and N; 2 women were designated as F and G. During running experiments, we gave instructions to each driver to observe traffic rules and speed limits based on as pause, as stipulated in the Road Traffic Law.

A. Correlation between driving behaviors and driving load sensitivity

Our previous study [9] specifically examined time-series variation of eye-gaze movements and face orientations before and after encountering a near-miss. Consequently, we extracted behavioral patterns characterizing the "distraction state" on driving, and attempted to derive the effective findings for engineering modeling. In this section, narrowing down the time axis on the sudden appearance of the bicycle from microscopic points of view, we should analyze the temporal relation between eye-gaze movements and face orientations from the start point of viewing to the end point of tracking with respect to the sudden appearance of the bicycle. Figure 4 depicts scatter diagrams of eye-gaze movements and face orientations from the starting to the end point of tracking in the "concentration state." Comparison to the scatter diagram of face orientations shows that the scatter diagram of eye-gaze movements has been scattered widely. We recognized a tendency to control eye-gaze movements before face orientations when searching for the gazing target. Therefore, we subdivide the time axis from the start point of viewing to the end point of tracking, and analyze the eye-gaze movements and the face orientations by particularly addressing the timing to be shown on the scatter diagram. Figure 5 depicts the results by which only the eye-gaze movements appeared to scatter diagrams ahead. By contrast, Figure 6 presents results for which both eye-gaze movements and face orientations appeared simultaneously in scatter diagrams. In Figures 5 and 6, respectively, (a) is a scene image sudden appearance of the bicycle, (b) a scatter diagram of eye-gaze movements, and (c) is a scatter diagram of face orientations. As understood from Figure 5, when the driver was able to view the sudden appearance of the bicycle at first, we were unable to observe changes in the scatter diagram of face orientations. Subsequently, we confirmed the scatter diagram in which the eye-gaze movements and the face orientations were matched in Figure 6. Therefore, we believe that the driver had to view and track the sudden appearance of the bicycle for the first time at this timing.

Next, we consider the effect of the driving load sensitivity to eye-gaze behaviors in the "distraction state" and the "concentration state." In this study, particularly addressing

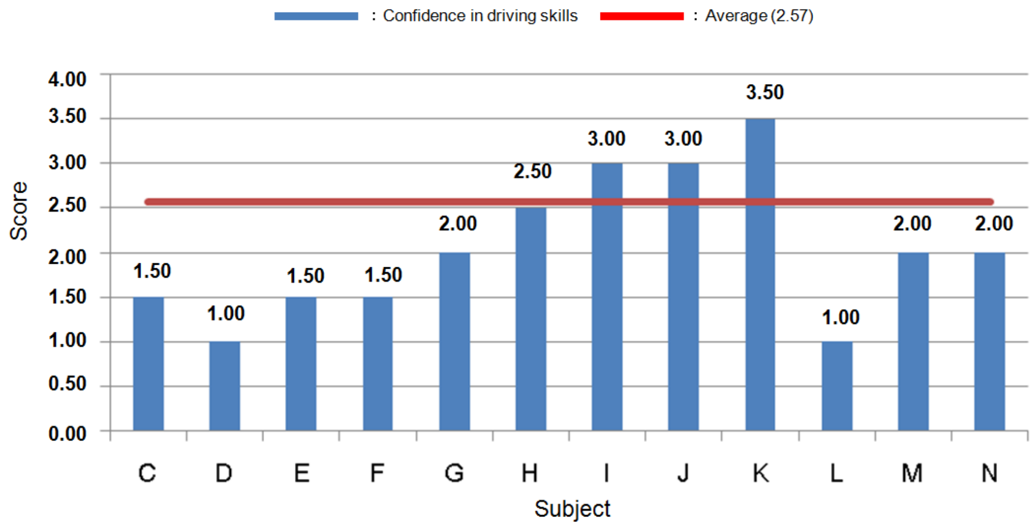


Figure 8. Calculation results of "confidence in driving skills" for all participants.

the "concentration inhibition for driving" from the 10 indices of the driving load sensitivity, Figure 7 depicts the result of whole participants. The value of the driving load sensitivity is calculated at 0.50 steps from 1.00 to 5.00, the average value (i.e., 2.63) in Figure 7, which indicates the survey result intended for the drivers of 20 years old to 74 years old conducted by Research Institute of Human Engineering for Quality Life. In the "concentration inhibition for driving" by performing mental arithmetic tasks and driving tasks simultaneously, we have classified "easily burdened group (i.e., C, E, F, G, I, K, L, and M)" and "not easily burdened group (i.e., D, H, J, and N)." Then, we analyzed gazing behaviors using heat maps of eye-gaze movements. Results show that for the easily burdened group the "concentration inhibition for driving" tended to fall into a "distracted state" by performing mental arithmetic tasks and driving tasks simultaneously. We inferred that the "easily burdened group" had unintentionally devoted attention resources to mental arithmetic tasks because their eye-gaze movements were distributed when approaching intersections.

Additionally, we consider the effect of the driving style to eye-gaze behaviors in the "distraction state" and the "concentration state." Particularly addressing the "confidence in driving skills" from the 9 indices of the driving style, Figure 8 depicts the result of whole participants. The value is calculated at 0.50 steps from 1.00 to 4.00, the average value (i.e., 2.57) in Figure 8, which indicates the survey result as same as the driving load sensitivity. In the easily burdened group (i.e., C, E, F, G, I, K, L, and M), we have confirmed subjects I and K having with the "confidence in driving skills." On the other hand, subjects C, E, F, G, L, and M don't have the "confidence in driving skills." We have expected that a concentration state and distraction state strongly depend on both "concentration inhibition for driving" and "confidence in driving skills" of each driver.

B. *Modeling of internal states estimation*

BNs can be used as a probabilistic model to visualize complicated dependence of the target problem by a graph structure. They represent dependence between variables by non-circular effective links. BNs represent random variables by nodes. Furthermore, BNs combine nodes with directed links and define the mutual dependence of variables as probability distributions. The node under the directed link is called the parent node. The node at the end of the directed link is called the child node. A random variable we want to find is defined as X. The value of the observed variable is defined as e. A normalization constant is defined as π. The probability of propagation from parent node or child node is defined as π and λ. An arbitrary *a posteriori* probability is calculable locally using the following equation: by stochastic inference using BNs, the probability distribution of the target random variable is obtainable.

$$P(X_j|e) = \alpha\lambda(X_i)\pi(X_j) \qquad (1)$$

The internal state estimation model handled in this research consists of 8 nodes: 2 nodes corresponding to the face orientation and eye-gaze movement of body information (Head and Eye), 1 node corresponding to the heart rate of the biometric information (Heart), 3 nodes corresponding to handles, accelerator, and brake of operation information (Handle, Accelerator, Brake), and 2 nodes corresponding to the driving scene and the driver's internal state (Scene, State). Table I presents the list of nodes. For each node from "Head" to "Brake," changes of values were recorded every 3 s because the mental arithmetic task given every 3 s in the driving experiment. "Scene" is an assigned number denoting the state according to the three driving scenes (straight line, intersection entry, and right turn) from characteristics of the simulation course.

For building the internal state estimation model, we use BAYONET, which is Bayesian Network building support software by NTT Data Matrix Systems. BAYONET has the function of learning the characteristics of the target dataset and automatically constructing a network. Now, based on the precondition that the driver's internal state and operation information are influenced according to the driving scene, the driving scene "Scene" is set as the parent node. Then the internal state "State" and the operation information "Handle, " "Accelerator," and "Brake" are set as child nodes. Then the directed link is set manually. Furthermore, based on the hypothesis that the driver's internal state is likely to appear in the physical information and in the biological information, the physical information "Head," "Eye" and the biological information "Heart" are set as parent nodes. The internal state "State" is set as a child node. The directed link is set manually.

We consider that the driver's internal state "State" is determined by the distribution of attention resources accompanying the performance of the driving task and the mental arithmetic task. Therefore, we quantify the distribution of attention resources for driving tasks and mental arithmetic tasks, and define the driver's internal state using these values. We use the number of occurrences of saccades every 3 s as the evaluation value of the attention resource distribution to driving tasks. Because saccades are a rapid movement of the eye-gaze movement from the gaze point to the gaze point, many saccades occurred during driving while devoting attention to the surrounding environment. The evaluation value of i-th driving task D_i is defined using the following equation. The number of occurrences of saccade S_i is normalized as a value of 0–1.

$$D_i = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}} \quad (2)$$

We use the time from the instant the question is presented to the end of the answer as the evaluation value of attention resource distribution to mental arithmetic tasks. A quicker answer reflects concentration on the mental arithmetic task and represents allocation of more attention resources. Therefore, the evaluation value is higher. However, in the case of no answer or a wrong answer, we set the evaluation value as the lowest because attention resources are not allocated to the mental arithmetic task. Specifically, it was set to the same value as the longest response time of each subject. The evaluation value of the mental arithmetic task at i-th point A_i is defined as shown in the following equation. In addition, T_i, which is the time from the question entry to the end of the reply, was normalized as a value between 0 and 1. A_i is set to approach 1 for rapid answers.

TABLE I. DEFINITION OF BNs NODES

Node	Contents
HEAD	Displacement value of head pose movements for 3 s [m]
EYE	Displacement value of eye-gaze movements for 3 s [m]
HEART	Variation of instantaneous heart rate for 3 s [s^{-1}]
HANDLE	Operation value of steering for 3 s (-1 ~ +1)
ACCEL	Operation value of accelerator for 3 s (-1 ~ +1)
BRAKE	Operation value of brake for 3 s (-1 ~ +1)
SCENE	Driving scenes (1: straight, 2: approaching intersections, 3: turn right)
STATE	Internal states (1: concentration state, 2: neutral state, 3: distraction state)

$$A_i = 1 - \frac{T_i - T_{\min}}{T_{\max} - T_{\min}} \quad (3)$$

From the above, the *i*-th evaluation value of the driver's internal state $State_i$ is defined from the evaluation value of the driving task D_i and the evaluation value of the mental arithmetic task A_i using the following equation.

$$State_i = \frac{D_i + A_i}{2} \quad (4)$$

C. *Analysis of internal states estimation model*

The state in which attention resources are not properly allocated according to the driving scene is called a distraction state. If a distraction state continues, then it engenders dangerous driving and a higher probability of causing an accident. Additionally, it is expected that a concentration state and distraction state defined as the driver's internal state depend on the driving skill of each driver. For this reason, we think that an analysis particularly addressing the driving style and driving workload sensitivity based on check sheets is important. As described herein, we specifically examine subject C, who has less opportunity to drive, and subject I, who drives on a daily basis. Figure 9 presents test results related to the driving style. Figure 10 shows test results reflecting the driving workload sensitivity. Average values shown in Figure 9 and Figure 10 are the values of results from about 540 men and women for a survey conducted by the Research Institute of Human Engineering for Quality of Life (HQL). In the driving style of Figure 9, subject C has high "passivity to driving" and "a propensity for anxiety." In addition, Subject I had a high tendency for "confidence in driving skills" and a "car as a status symbol." From this, they are found to have polar opposite driving styles. In terms of the driving workload sensitivity in Figure 10, both subjects had a high "concentration inhibition for driving."

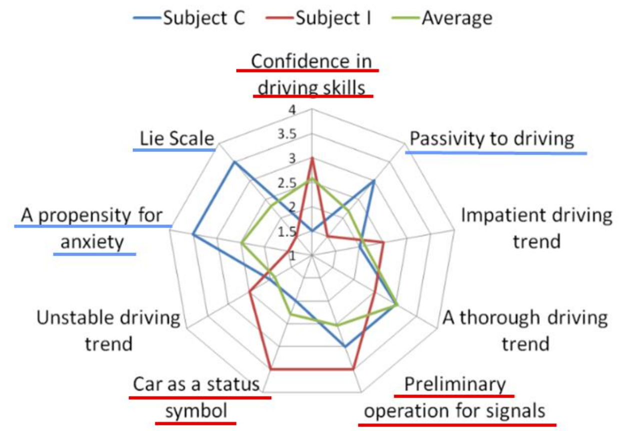


Figure 9. Driving style results of subjects C and I.

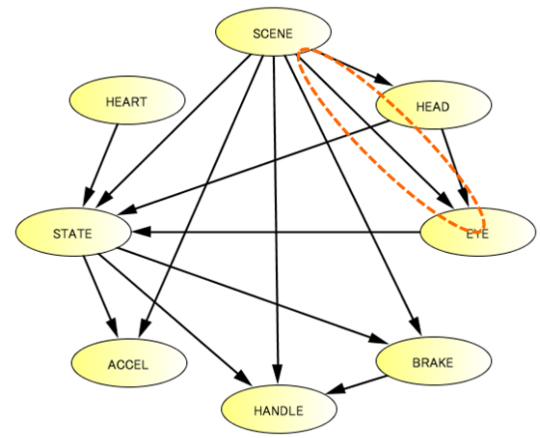


Figure 11. Internal state estimation model for risky driving (Subject C).

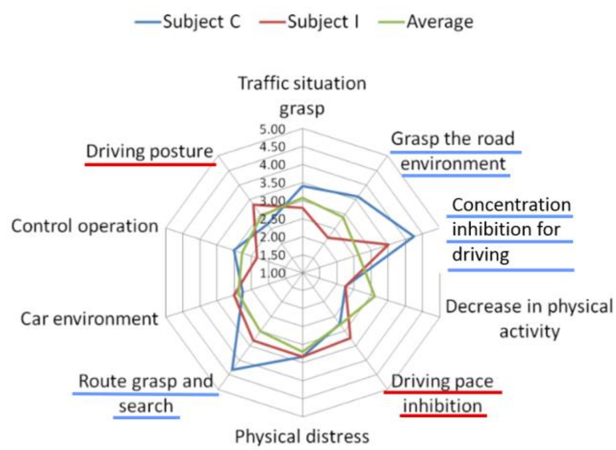


Figure 10. Driving workload sensitivity results of subjects C and I.

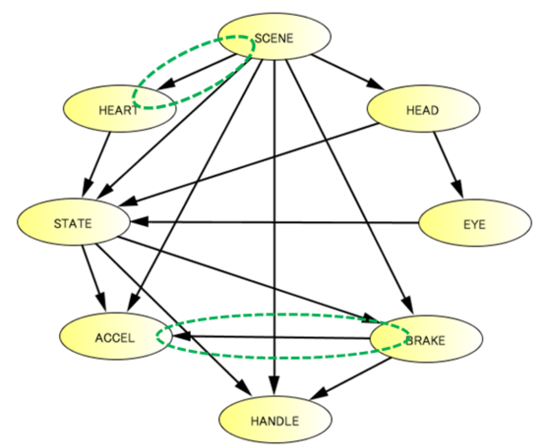
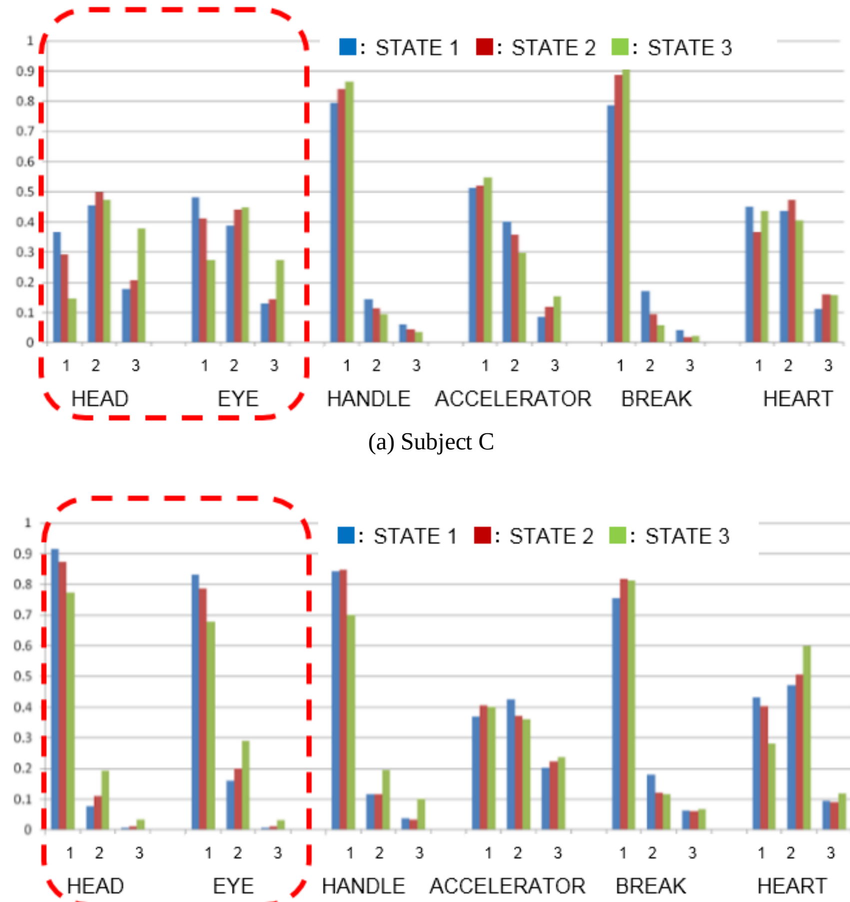


Figure 12. Internal state estimation model for risky driving (Subject I).

Subject I exhibited a tendency to have a higher "driving pace inhibition." In other words, we estimate that both subjects have characteristics that easily fall into a distraction state because of simultaneous performance of the driving task and mental arithmetic task.

The learning data used to construct the internal state estimation model are a joining of all data acquired in the four driving experiments conducted with different weather and time zones. The number of data of subject C was 634. The number of data of subject I was 619. In addition, these data were discretized in three stages using K-means method. The model of subject C is presented in Figure 11. The model of subject I is shown in Figure 12. Comparing Figure 11 with Figure 12 enables confirmation of the existence of a directed link unique to each subject. Because the directed links connect the nodes in a form optimized to raise the accuracy of the inference, significant correlation exists between the nodes to which the directed links are established. For example, in Subject C in Figure 11, a directed link from the "Scene" node to the "Eye" node is established, but in subject I in Figure 12, its directed link is not established. In other words, subject C's eye-gaze movement is more likely to be affected depending on the driving scene. Conversely, subject I's eye-gaze movement might be affected only slightly. In subject I in Figure 12, a directed link is established from the "Scene" node to the "Heart" node and from the "Brake" node to the "Accelerator" node, respectively, but in the subject C in Figure 11, its directed link is not established. Particularly because the directed link from the "Brake" node to the "Accelerator" node shows a clear causal relation between the brake and the accelerator operation amount, it can be regarded as the skill level of the driving operation. The feature shown above demonstrates that it is possible to construct an internal state estimation model that is unique to the subject because of differences in driving style and driving workload sensitivity.



(a) Subject C

(b) Subject I

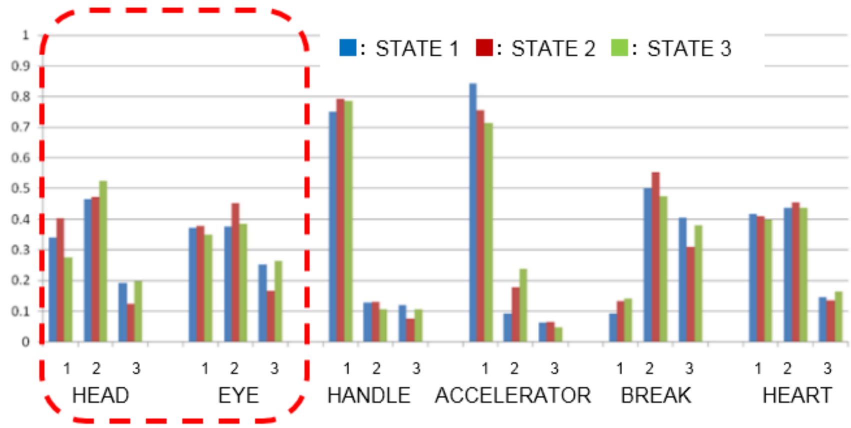
Figure 13. Probability distribution of respective nodes with evidence of the straight running period.

D. *Stochastic reasoning by internal states estimation model*

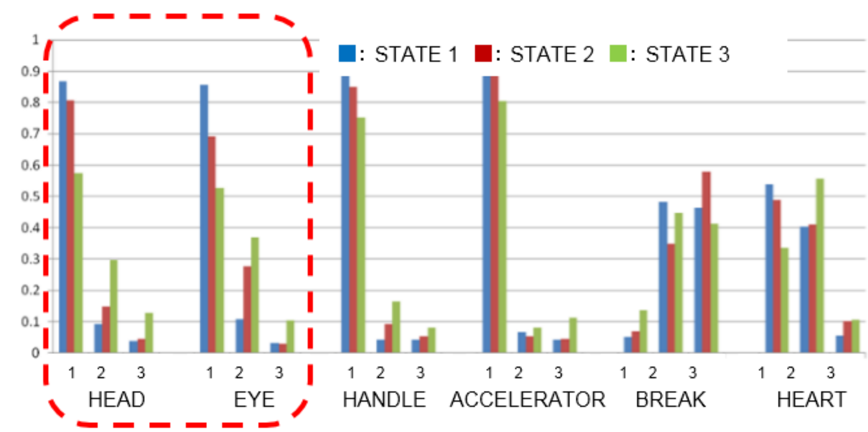
Next, we identify characteristic behavior patterns corresponding to the three driving scenes and analyze the correlation between physical information and operation information, each of which is likely to be affected by the distracted state using the constructed internal state estimation model. The three driving scenes are controlled by evidence of the "Scene" node. Scenes are classified into a straight section "Scene 1," an entrance section "Scene 2," and a right turn section "Scene 3." Subsequently, we gave evidence to the ‘State’ node representing the driver’s internal state for each driving scene, and compared the probability distribution of each node in the concentrating state "State 3" and the distracted state "State 1." Figures 13–15 present the posterior probability distributions of subjects C and I obtained by stochastic reasoning corresponding to three driving scenes. In the figure, values 1–3 indicated by each node on the abscissa show the magnitude of the movement amount and the operation amount. The ordinate axis shows values of their posterior probabilities.

In the straight section of Figure 13 and the approach section of Figure 14, devoting attention to the posterior probability distribution of the ‘Head’ node and the "Eye" node, subject C has a large face orientation and movement amount of the eye-gaze movement. Conversely, subject I's movement amount is small. Regarding the driving style, subject C has high "passivity to driving" and "a propensity for anxiety." Subject I has a high "confidence in driving skill," so it can be inferred that this is attributable to the difference in driving skill levels. In addition, in the right turn section of Figure 15, a tendency is apparent by which the face orientation and eye-gaze movement amount of subject I become greater than in the straight section of Figure 13. In other words, it can be inferred that subject I was driving while identifying points to be noticed according to each driving scene, in contrast to subject C, who did not.

In addition, when comparing concentrating state "State 3" and distracted state "State 1" in a distracted state in both driving scenes of both subjects, the amounts of movement



(a) Subject C



(b) Subject I

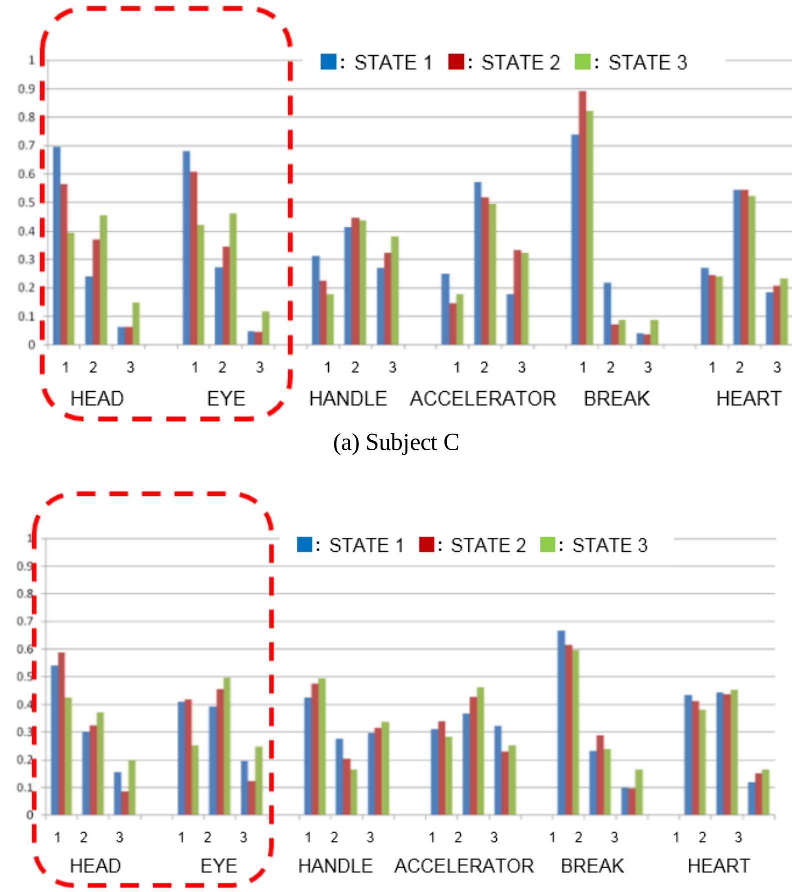
Figure 14. Probability distribution of respective nodes with evidence of approaching the intersection period.

of the face orientation and the eye-gaze movement are small (posterior probability of "Head 1" and "Eye 1": distracted state > concentrating state). However, in the concentrating state, for both driving scenes of both subjects, the amount of movement the face orientation and the eye-gaze movement is large (posterior probability of "Head 3" and "Eye 3": distracted state > concentrating state). These tendencies are regarded as evident in the change in the amount of the movement of the face orientation and the eye-gaze movement for both subjects as a result of depriving attention resources by the simultaneous execution of the driving task and the mental arithmetic task. Next, devoting attention to the posterior probability distribution of the 'Handle' node in the straight section of Figure 13 and the entry section of Figure 14, in the distracted state, the steering wheel operation amount of subject C is increased, whereas the steering wheel operation amount of subject I is decreased. Moreover, subject I, more than subject C, showed a difference in posterior probability when the concentrating state changed to the distracted state. This tendency suggests that the simultaneous execution of mental arithmetic tasks is strongly influential because subject I shows a burden from "distributing the concentration of the driving" and "driving pace inhibition."

Finally, we specifically examine the probability distribution of the "Heart" node of the instantaneous heart rate. In the internal state estimation model of subject I presented in Figure 12, a direct link is established from the "Scene" node to the "Heart" node. For subject C presented in Figure 10, the directed link is not recognized. By devoting attention to this point and by comparing the posterior probability distributions of both subjects shown in Figures 13–15, subject I presents a marked difference between the posterior probability of the concentrating state and distracted state in each driving scene, but subject C shows no marked difference. This result is attributable to the presence or absence of an effective link in the internal state estimation model that is unique to each subject.

E. Validation of internal states estimation model

After constructing an internal states estimation model with three datasets of learning data out of the four running



(a) Subject C

(b) Subject I

Figure 15. Probability distribution of respective nodes with evidence of turning to the right period.

experiments conducted with different weather and time zones (sunny day / rainy day / sunny night / rainy night), probabilistic inference was conducted for the model using the remaining 1 dataset as test data. Then the accuracy of the model was verified through cross validation. Table II presents the results. Assuming that the number of verification data is *N* and that the correct number of inference results is *C*, then the correct answer rate is represented as *C* / *N*. Therefore, the average value was calculated and inferred by all combinations (model construction and test of four patterns). It is noteworthy that the "Scene" node and the "State" node were used as explanatory variables. The other six nodes were set as objective variables to find the correct answer rate.

As shown in Table II, particularly addressing Subject I, it indicates a high correct answer rate in all the nodes except the "Heart" node. Especially, in the "Eye", the "Head," the "Handle," and the "Brake" nodes directly connected to judgment / operation of driving behavior, the correct answer rate of more than 70% can be confirmed for each node. These numerical values show the possibility of extracting and defining a characteristic driving behavior pattern when shifting from the driving concentration state to the distracted state. Results suggest effectiveness of the internal state estimation model that is unique to the subject. However, the correct answer rates of the "Eye" node and the "Head" node of subject C tended to be lower than those of subject I because subject I was driving without glasses, whereas subject C was wearing glasses while driving. Therefore, we infer that the difference in detection accuracy between the eye-gaze movement and the face orientation was influenced.

TABLE II. RESULTS OF ACCURACY VALIDATION

Node	Subject C	Subject I
HEAD	0.48715	0.779525
EYE	0.417975	0.707025
HEART	0.49720	0.443275
HANDLE	0.70335	0.746700
ACCEL	0.52265	0.574150
BRAKE	0.71645	0.755225

V. Conclusion and Future Work

For this study, we analyzed driver behavior changes to ascertain the time at which a driver becomes distracted. In an earlier study, we conducted a driving experiment simulating a distraction state and obtained a dataset. Additionally, we constructed a Bayesian model incorporating the driver's internal state as a node for two subjects, and executed probabilistic reasoning. We analyzed the relation between inference results and driving style / driving workload sensitivity. Results clarified the following points. Dividing "watching behaviors" when approaching the intersection and the "safety confirmation behaviors" after a temporary stop is effective for analyses of the behavioral patterns characterizing the "distraction state." It is possible to construct a specific internal state estimation model of the subject by differences in the driving style and driving workload sensitivity. In addition, results show a unique directed link of the subject. Although we analyzed only two subjects in the present study, we expect to evaluate more subjects and analysis items in future research. Additionally, it is considered that temporal information is necessary for dealing with driving behavior. Therefore, we plan to conduct research with modeling using DBNs.

ACKNOWLEDGMENT

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 25330325.

References

[1] K. Sato, M. Sawataishi, H. Madokoro, M. Ito, and S. Kadowaki, "Basic Study for Distraction State Detection based on Driver Body Information Analysis," The twelfth International Multi-Conference on Computing in the Global Information Technology (ICCGI 2017), July 2017.

[2] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining Driver Visual Attention with One Camera," IEEE Trans. on Intelligent Transportation Systems, Vol. 4, No. 4, pp. 205-218, 2003.

[3] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: a review," IEEE Trans. on Intelligent Transportation Systems, Vol. 12, No. 2, pp. 596-614, 2011.

[4] Y. Tanaka and T. Bando, "Anomaly Driver State Detection from Vehicle Behavior Data," Society of Automotive Engineers of Japan, Inc. (JSAE), Vol. 44, No. 2, pp. 685-690, 2013.

[5] K. Kircher, C. Ahlstrom, and A. Kircher, "Comparison of Two Eye-gaze based on Real-time Driver Distraction Detection Algorithms in a Small-scale Field Operational Test," Proc. Fifth Int. Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 16-23, 2009.

[6] R. Yonetani, H. Kawashima, T. Hirayama, and T. Matsuyama, "Mental Focus Analysis using the Spatiotemporal Correlation between Visual Saliency and Eye Movements," Journal of Information Proc., Vol. 20, No. 1, pp. 267-276, 2012.

[7] A. Tsuchida, H. Kawanaka, and K. Oguri, "Classification of Driver's Drowsy States with Respect to Individual Difference Based on Application of Subjective Rating by Facial Expression," The Institute of Electronics, Information, and Communication Engineers, Technical Report, ITS Vol. 111, No. 441, pp. 251-256, 2012-02-13.

[8] T. Omi, "Driver Sleepiness Detection by Video Image Processing, Japanese Society for Artificial Organs," Vol. 42, No. 1, pp. 99-103, 2013. doi: 10.11392/jsao.42.99.

[9] K. Sato, M. Ito, H. Madokoro, and S. Kadowaki, "Driver Body Information Analysis for Distraction State Detection," International Conference on Vehicular Electronics and Safety, pp. 13-18, 2015.

[10] M. Sawataishi, K. Sato, H. Madokoro, M. Ito, and S. Kadowaki, "Driver Internal State Estimative Model for Distracted State Detection," IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2504-2509, 2017.

[11] Y. Motomura, "Probabilistic reasoning algorithms and their experiments in Bayesian network," IEICE, pp. 157-162, 2004.

[12] M. Miyaji, H. Kawanaka, and Oguri, "Driver's Cognitive Distraction Detection using Physiological Features by the AdaBoost," 12th International IEEE Conference on Intelligent Transportation Systems, pp. 1-6, 2009.

[13] T. Hirayama, K. Mase, and K. Takeda, "Timing Analysis of Driver Gaze under Cognitive Distraction toward Peripheral Vehicle Behavior," The 26th Annual Conference of the Japanese Society for Artificial Intelligence, pp. 1-4, 2012.

[14] T. Hirayama, S. Sato, K. Mase, C. Miyajima, and K. Takeda, "Analysis of peripheral vehicular behavior in driver's gaze transition: Differences between driver's neutral and cognitive distraction states," 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 962-967, 2014.

[15] G. Abe, K. Kikuchi, R. Iwaki, and T. Fujii, "Effects of Cognitive Distraction on Driver's Visual Attention," The Society of Mechanical Engineers (JSME), Transactions of the JSME (c), Vol. 76, No. 767, pp. 1662-1668, 2010.

[16] R. Honma, G. Abe, and K. Kikuchi, "Characteristics of Visual Attention while Driving under the State of Drowsiness," Society of Automotive Engineers of Japan, Inc. (JSAE), Vol. 42, No. 5, pp. 1217-1222, 2011.

[17] S. Kumano, "Sensing Human Actions for Mental State Estimation," University of Tokyo, 2009. http://hdl.handle.net/2261/28031 [retrieved: May, 2018]

[18] K. Kircher and C. Ahlstrom, "The Driver Distraction Detection Algorithm AttenD," in Driver Distraction and Inattention: Advances in Research and Countermeasures, M. Regan, J. D. Lee, and T. W. Victor, Eds. Surrey, U.K.: Ashgate, pp. 327-348, 2012.

[19] C. Ahlstrom, K. Kircher, and A. Kircher, "A Gaze-Based Driver Distraction Warning System and Its Effect on Visual Behavior," IEEE Transactions on Intelligent Transportation Systems, vol.14, no.2, pp. 965-973, 2013.

[20] M. Ishibashi, M. Okuwa, S. Doi, and M. Akamatsu, "HQL Driving Style Questionnaire: DSQ," Research Institute of Human Engineering for Quality Life, 2003.

[21] M. Ishibashi, M. Okuwa, S. Doi, and M. Akamatsu, "HQL Workload Sensitivity Questionnaire: WSQ," Research Institute of Human Engineering for Quality Life, 2003.

N-Back Training and Transfer Effects in Healthy Young and Older Subjects Gauged Using EEG: A Preliminary Study

Valentina Pergher, Benjamin Wittevrongel
KU Leuven - University of Leuven, Department of Neurosciences,
Laboratory for Neuro- & Psychophysiology,
Leuven, Belgium
email: valentina.pergher@kuleuven.be,
benjamin.wittevrongel@kuleuven.be

Birgitte Schoenmakers
Academic Centre of General Practice,
KU Leuven - University of Leuven,
Leuven, Belgium
email: birgitte.schoenmakers@med.kuleuven.be

Jos Tournoy
Faculty of Medicine, Department of Experimental Medicine,
Division of Gerontology and Geriatrics,
KU Leuven - University of Leuven ,
Leuven, Belgium
email: jos.tournoy@uzleuven.be

John Arsenault, Marc M. Van Hulle
KU Leuven - University of Leuven, Department of Neurosciences,
Laboratory for Neuro- & Psychophysiology,
Leuven, Belgium
email: john.arsenault@kuleuven.be,
marc.vanhulle@med.kuleuven.be

***Abstract*- Executive function performance of older individuals is lower compared to young adults. We investigate whether N-Back working memory (WM) training improves both trained WM- and untrained cognitive function performance (transfer effects). As previous studies showed that electroencephalogram (EEG) responses, in particular Event Related Potentials (ERPs), vary with task difficulty level and age, we focused on the relation between ERPs-P300 and task difficulty level in young and older adults. We used two groups of healthy young and older participants to assess the effect of N-Back training: cognitive training group (CTG) and passive control group (PCG). CTG performed an N-Back task with 3 difficulty levels (1, 2, 3-Back), and PCG did not undergo any training. Pre- and post-tests were administered to both groups to gauge any transfer effects (spatial memory, attention and fluid intelligence). Our results show age-related differences in P300-ERPs, reaction time and accuracy for N-Back training task and in spatial memory and reasoning pre post-tests. Improvements in the trained task are stronger for CTG young than CTG older individuals. Furthermore, transfer effects to attention (TOVA) were found in both young and older adults for CTG, showing the benefits of the training.**

Keywords-EEG; working memory training; transfer effects; P300 ERP; young and older adults.

I. INTRODUCTION

Cognitive training is a powerful tool to explore the plasticity of the aging brain to improve cognitive functions such as intelligence, episodic memory, working memory, and executive functions [1][2][3]. Morrison and Chein [4], in a review of working memory (WM) training, divided the training approaches in two categories: strategy training (domain specific strategies), and core training (repetition of demanding WM tasks). In our study, we adhered to the second approach, where the participants were exposed several times to a repetition of visual stimuli during a WM task.

Working memory (WM), as defined by Baddeley [2], refers to the temporary storage and manipulation of information necessary to execute complex cognitive tasks. WM training was originally used to enhance WM in neuropsychiatric subjects with a WM deficit, such as attention deficit hyperactivity disorder (ADHD) [5] and several authors studied the mechanisms behind and the effect of WM training [6][7].

The N-back task is a working memory task introduced by Wayne Kirchner in 1958 [8] as a visuo-spatial task with four load factors (‘‘0-Back’’ to ‘‘3-Back’’), and by Mackworth [9] as a visual letter task with up to six load factors. Gevins et al. [10] introduced it into neuroscience by using it as a ‘‘visuomotor memory task’’ with one load factor (3-Back). The N-back task involves multiple processes and is considered a dual task: working memory updating, which involves the encoding of incoming stimuli, the monitoring, maintenance, and updating of the sequence, and stimulus matching (matching the current stimulus to the one that occurred N positions back in the sequence). It reflects a number of core Executive Functions (EFs) besides working memory, such as inhibitory control and cognitive flexibility, as well as other higher-order EFs, such as problem solving, decision making, selective attention, among other functions [11]. The N-Back task requires participants to maintain stimulus information necessary for successful task performance in working memory across multiple trials [11]. It has been shown that the N-Back task consistently activates dorsolateral prefrontal cortex (DLPFC), as well as parietal regions in the adult brain [12]. Schneiders et al. [13] have shown that, using N-Back training, it is possible to achieve an improvement in task performance and an alteration in brain activity, such as a decreased activation in the right superior middle frontal gyrus (BA 6) and posterior parietal regions (BA 40).

Following a series of studies, Jaeggi et al. [14][15] reported that by performing an N-Back task, the effects of WM training transfer to untrained tasks requiring WM (transfer effects) improve upon a complex human ability known as fluid intelligence. The findings of Jaeggi et al. [14] also support the hypothesis that transfer effects to general cognitive functions can be achieved after an N-Back training for tasks that conceptually overlap, albeit only slightly, with the N-Back task. Training of the general fronto-parietal WM network should lead to improvements in cognitive functions that rely on the same network [10]. This general overlap hypothesis predicts that if training considerably engages the fronto-parietal WM network and the transfer task generates a similar activation pattern, an extensive training of this network will yield a general boosting of cognitive functions. An alternative hypothesis predicts that WM training effects transfer only if training improves specific cognitive processes required in both training and transfer tasks. Dahlin et al. [16] found transfer, after WM updating training, to an N-Back task that resembled the original trained task in also relying on updating processes, but not to a Stroop task that involved inhibition but no updating. Furthermore, Dahlin et al. [16] and Li et al. [2] reported that training on an N-Back task improves WM training for both young and older healthy subjects. Interestingly, Dahlin et al. [16] found larger improvements in young healthy subjects, whereas Li et al. [2] found comparable improvements in both groups. The reasons for this difference could be the frequency and duration of the training, and the overlap of the elements of a skill between the trained and the specific transfer task. Despite this difference, both studies demonstrate working memory plasticity in young and older adults, and the possibility to obtain near-transfer effects that are stable in time. So, although the degree of plasticity varies across studies, the potential of the brain to reorganize in response to demands is found in people of all ages [17][18][19].

The aim of our study was to verify whether N-Back task performance improves after N-Back training, and whether transfer effects to other (untrained) cognitive functions are obtained, such as in spatial memory, attention and reasoning, in two different groups of healthy young and older subjects: cognitive training group (CTG) and passive control group (PCG). We recorded EEG responses during all training sessions, and focused on the P300, an ERP component related to working memory [20].

The paper is organized as follows. In Section II, we describe the material and methods (subjects, procedure, EEG recording). In Section III, we focus on the behavioral and P300-ERPs results using WM training and the transfer effects pre- and post-training in young and older adults. Finally, in Section IV, we discuss our results and propose a number of technical and conceptual goals for future studies.

II. MATERIALS AND METHODS

In this section, we describe the participant recruitment, training procedure and EEG recording.

A. *Subjects*

We recruited 17 healthy young subjects (9 females, 8 males, mean age 29 years, range 24-34 years), undergraduate or graduate students from KU Leuven, and 19 healthy older subjects (13 females, 6 males, mean age 62.25 years old, range between 50-69 years old) via posters, social media, and KU Leuven's Academic Center for General Practice. Participants were healthy, with normal or corrected vision, without any history of psychiatric or neurological diseases, not on any medication, and never participated in working memory training.

B. *Cognitive training*

Participants were assigned to two sub-groups, cognitive training group (N=9, young; N=10, older) and passive control group (N=8, young; N=9, older), to evaluate improvements in task performance after WM training and to record any transfer effects to other cognitive tasks (see further for their definition). During all training sessions, EEG was recorded (see also further). For the cognitive training group, all subjects (i.e., young and older ones) performed WM training with visual feedback of the correctness of their behavioral responses and received a monetary reward (with a maximum of 10 € if all responses were correct). The control group did not undergo any training.

C. *Transfer effects*

A battery of cognitive tests were administered before and after training (pre and post-tests, note that for the control group there was no training between tests) to see if there were transfer effects to attention, spatial memory, reasoning, and intelligence. The study was approved by our university's ethical committee and when subjects agreed to participate in the experiment they signed the informed consent form prior approved by the said committee.

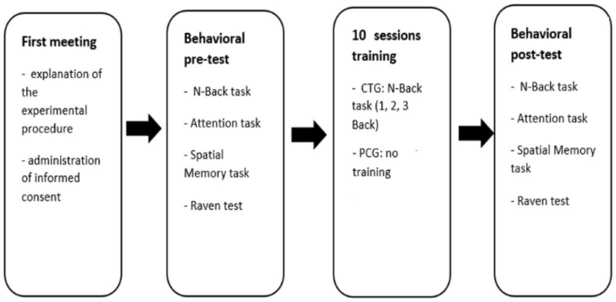


Figure 1. Study design

D. *Procedure*

Subjects participated in an N-back task in which a sequence of stimuli was shown and the task was to decide whether the current stimulus matched one shown N items earlier (Figure 2).

The stimuli were presented for 1000ms followed by a 2000ms Inter-stimulus interval (ISI), jittered by ± 100 ms, during which the picture was replaced by a fixation cross. This is the moment when participants needed to press a button when the stimulus was a target. We had 33% of our pictures as targets.

Sequences with identical difficulty levels (1-back, 2-back, 3-back) were grouped into 2 min. blocks across four sessions. Each session included two repetitions of 3 sequences. In total there were 8 blocks. For each sequence, there were 60 stimuli presented in pseudorandom order. Before starting with the first three sequences, a training session consisting of ten stimuli for each difficulty level was used to explain the N-Back task. Subjects performed an N-Back training during 10 sessions, 3 times per week (30 minutes each time), as shown in Figure 1. This is in line with literature reports on significant training and transfer effects obtained after 3 weeks of training [14][21-24].

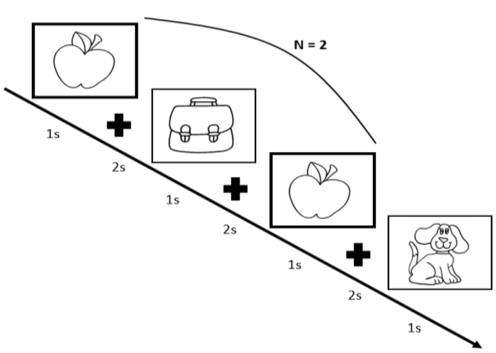


Figure 2. Graphical rendition of 2(N)-back task

All participants were administered a battery of pre- and post-tests to evaluate whether there were transfer effects to other cognitive functions. We used Test of Variables of Attention (TOVA) [25], Spatial Working Memory Test (CORSI) [26] and RAVEN, a fluid intelligence test [27]. The behavioral pre- and post-tests were administered to compare task performance between groups (cognitive training, active control and passive control groups) in the trained (N-Back task) and untrained tasks (TOVA, CORSI and RAVEN test).

N-back task and transfer tasks had similarities and differences [14][27][28]. The spatial memory task (Corsi test) engaged WM updating processes just as the N-Back task, but differed in stimuli (squares in Corsi task vs. pictures in the N-Back task) and task rules (recognition of previously presented items in the N-Back tasks vs. recollection of items in the updating transfer tasks). Given these similarities and differences, we are using near transfer tasks according to Karbach and Kray [29].

In the first experimental session (pre-test), each participant was informed about the experimental procedure and invited to sign the informed consent form. The day after the first meeting, the participants performed the behavioral pre-test session, and from the third meeting, the training groups (CTG) started the training procedure of CTG. At the beginning of each training session, an EOG calibration session was performed to model eye movements and blinks using the AAA method described in Croft and Barry [30].

E. *EEG recording*

EEG was recorded continuously from 32 Ag/AgCl electrodes at a sampling rate of 2 kHz using a SynampsRT device (Neuroscan, Australia). The electrodes were placed at O1,Oz, O2, PO3, PO4, P8, P4, Pz, P3, P7, TP9, CP5, CP1, CP2, CP6, TP10, T7, C3, Cz, C4, T8, FC6, FC2, FC1, FC5, F3, Fz, F4, AF3, AF4, Fp1, Fp2. The reference was placed at AFz and the ground at CPz. Additionally, four electrodes were placed around the eyes, on the upper and lower side of

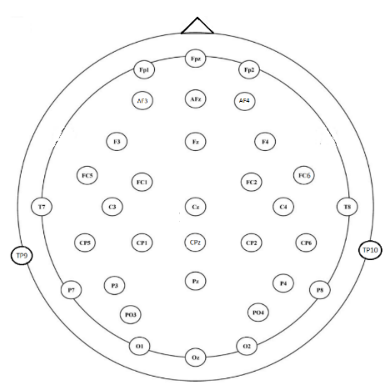


Figure 3. Names and distribution of electrodes.

the left eye (vertical) and near the external canthus of each eye (horizontal), for electro-oculogram recording (EOG, bi-polar recording).

The recorded EEG signal was re-referenced offline from the original reference to the average of two mastoid electrodes (TP9 and TP10), corrected for eye movement and blinking artifacts [30], band-pass filtered in the range of 0.1–30 Hz, and cut into epochs starting 200 ms pre- till 1000 ms post- stimulus onset. Baseline correction is performed by subtracting the average of the 200 ms pre-stimulus onset activity from the 1000 ms post-stimulus onset activity. Finally, the epochs are downsampled to 100 Hz and stored for ERP detection.

Recorded epochs with incorrect button press responses were excluded from further analysis. In addition, epochs with EEG signals greater than 100mV were also excluded. A two-way ANOVA (factors: n-back X target) was performed on all sampled EEG time points between 250 ms to 400 ms. Bonferroni correction for multiple comparisons was used across all samples within this time window.

III. RESULTS

In this section, we describe working memory training (behavioral and ERPs results) and transfer effects.

A. *Working memory training (behavioral)*

We also analyzed changes due to cognitive training by examining behavioral data (accuracy, reaction time (RT)) of CTG during N-Back training (10 sessions) of healthy young and older subjects (Figures 4 and 5). The purpose is to test our second hypothesis: training can improve related cognitive function performance, and transfer to other cognitive functions as shown by significant effects in RT and response accuracy.

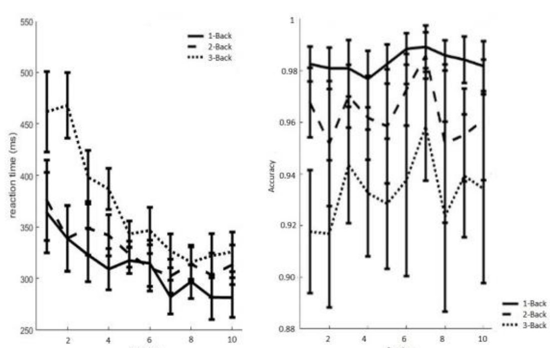


Figure 4. RT and accuracy during 10 sessions of cognitive training in CTG-young. Error bars indicate SEM (Standard error of the mean).

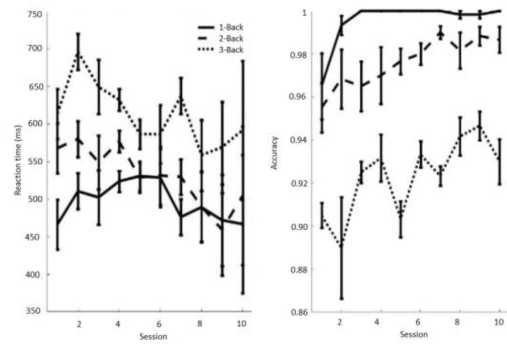


Figure 5. RT and accuracy during 10 sessions of cognitive training in CTG-older. Error bars indicate SEM.

For the CTG-young, we observed a reduction in RT as function of training sessions. To test this, we performed a three-way ANOVA across factors (N-back level, subject and session). We found a significant effect of session ($F_{(9)}=4.9$, $p<0.001$) confirming that RT indeed decreases with more training. Importantly, the N-Back level x session interaction was significant ($F_{(18)}=3.01$, $p<0.001$), which indicates that the N-back levels are differentially affected by training. In contrast, when we looked at accuracy, the main effect of session was not significant ($p=0.56$) indicating that accuracy did not substantially increase as a result of training although there was a main effect of N-back level confirming that task difficulty affected performance ($F_{(2)}=7.97$, $p<0.05$).

For the CTG-older, we also performed a three-way ANOVA across factors (N-back level, subject and session). We found for RT a significant effect of N-back level ($F_{(2)}=37.62$, $p<0.05$) indicating that task difficulty affects reaction time. For the accuracy we found significant effects for N-Back level ($F_{(2)}=119.58$, $p<0.001$) and session ($F_{(9)}=3.77$, $p<0.05$). Importantly, the N-Back level x session interaction was significant ($F_{(18)}=2.09$, $p<0.05$) indicating that accuracy improves with training and the gained improvements differ with N-Back level.

B. *Working memory training (ERPs results)*

Neuroimaging studies have shown that during N-Back task performance the most activated brain regions are the lateral premotor cortex, dorsal cingulate and medial premotor cortex, dorsolateral and ventrolateral prefrontal cortex, frontal poles, and medial and lateral posterior parietal cortex [10]. In addition, since several EEG studies showed that the midline electrodes are the most significant ones [31][32], we decided to analyze ERPs using electrodes Fz, Pz, and Cz. Data from mean P300 peak amplitude is presented in Figures 6 and 7 in three different moments (3 sessions/each moment) during training (first-, middle- and last sessions) for young and older adults.

P300 peak amplitude data from midline electrodes (Fz, Cz, Pz) were analyzed with a three-way ANOVA (N-Back level x target x session). P300 peak amplitude (target minus no-target) was higher for the N-Back levels that were easier (1 and 2-Back), and was lower for the more difficult one (3-Back), especially for healthy older subjects. P300 peak amplitude (difference between target and no-target) was largest for the frontal electrode (Fz) and decreased for the central (Cz) and posterior electrodes (Pz).

Significant differences were found in healthy young subjects for CTG (Figure 6) between first (1^{st}) and last (10^{th}) training session for 3-Back task in channel Fz ($F_{(1)}=6.4155$, $p<0.05$) and in channel Cz ($F_{(1)}=3.9479$, $p<0.05$), and between first and middle session (5^{th} session) for 3-Back task in channel Fz ($F_{(1)}=7.2620$, $p<0.01$), in channel Cz ($F_{(1)}=6.0811$, $p<0.05$), and in channel Pz ($F_{(1)}=5.4272$, $p<0.05$).

Furthermore, statistically, P300 amplitude between the 1^{st} and last session (10^{th}) showed higher amplitude for 3-Back task compared to the other N-Back difficulty levels (1 and 2-Back), indicating that young adults improved more for the most difficult task (3-Back). The P300 amplitude was largest for the frontal electrode (Fz) and decreased for the central (Cz) and posterior electrodes (Pz) in young adults for the difference between 1^{st} and 10^{th} session (Figure 6, right). We compared also 1^{st} and middle (5^{th}) session where we could already seen significant improvement for the most difficult task (3-Back), showing that 5 training sessions of N-Back training for healthy young adults could be sufficient to have significant improvement in the trained task. Taken together, these data support the observation that the P300 amplitude increase with training sessions, and that also for the 3-Back task, although the most difficult one, WM training was effective.

As to the healthy old subjects of CTG (CTG-old): we also analyzed the P300 amplitudes of the midline electrodes (Fz, Cz, Pz) with a three-way ANOVA (N-Back level x target x session) for CTG-old. We found significant differences between the first and the last session for 3-Back task in channel Pz ($F_{(1)}=6.2091$, $p<0.05$), showing that 3-Back task became easier for the participants. Furthermore, similar to the previous results for healthy young subjects, P300 amplitude in the first session was higher for the N-Back difficulty levels that were easier (1 and 2-Back), and lower for the more difficult one (3-Back). Also in this case, P300 amplitude (difference between target and non-target) was larger for the frontal electrode (Fz) than for the central (Cz) and posterior electrodes (Pz). After the training with older adults, the P300 became higher for the most difficult task (3-Back) in parietal region, showing that WM training can modify the amplitude (Figure 7). These findings confirm the results of Gevins et al. [35] who reported that training on an N-Back shows EEG changes in responses to changes in the mental effort required for task performance. We compared also 1^{st} and middle (5^{th}) session, but in contrast to young healthy adults, we did not find any significant improvement for the most difficult task (3-Back), showing that for healthy older subjects are necessary 10 training sessions of N-Back training. These data support the observation that the P300 amplitude increase with training sessions, and that also for the 3-Back task, although the most difficult one, WM training was effective.

Salminen et al. [36] showed after N-back training benefits for both young and older subjects in terms of behavioral response accuracies, with differences in improvement between young and older adults. Also Friedman, and Simpson [33] found changes in ERP amplitudes of young and older adults during oddball task performance. Given these observations, we looked for differences in P300 components of young and older healthy subjects. Our results showed significant effects for the interaction between age differences and target minus non-target for 2-Back task in channel Fz ($F_{(1)}=13.3309$, $p<0.001$), in channel Cz ($F_{(1)}= 6.4395$, $p<0.05$), and in channel Pz ($F_{(1)}=9.6903$, $p<0.01$).

C. *Transfer effects (Pre- and Post-tests)*

Means for each task are presented in Table I for the pre- and post-tests between young and older subjects. In Figures 8 and 9, a multivariate ANOVA (MANOVA) was conducted on intra and inter-groups (CTG and PCG young and older) and between sessions (pre- and post-tests). For young subjects, significant effects for accuracy in N-Back task between CTG and PCG ($F_{(1)}=6.21$, $p<0.05$), for pre- and post-testing, were observed. No significant differences in the other cognitive tests (CORSI and RAVEN) were found. For the N-Back task, significant effects were found for RT between CTG and PCG, for pre- and post-tests ($F_{(1)}=40.9$, $p<0.001$), for task difficulty level ($F_{(2)}=4.92$, $p<0.05$), for group x pre- and post-test interaction ($F_{(1)}=9.14$, $p<0.05$), and for pre- and post-test x N-Back level interaction ($F_{(2)}=3.54$, $p<0.05$).

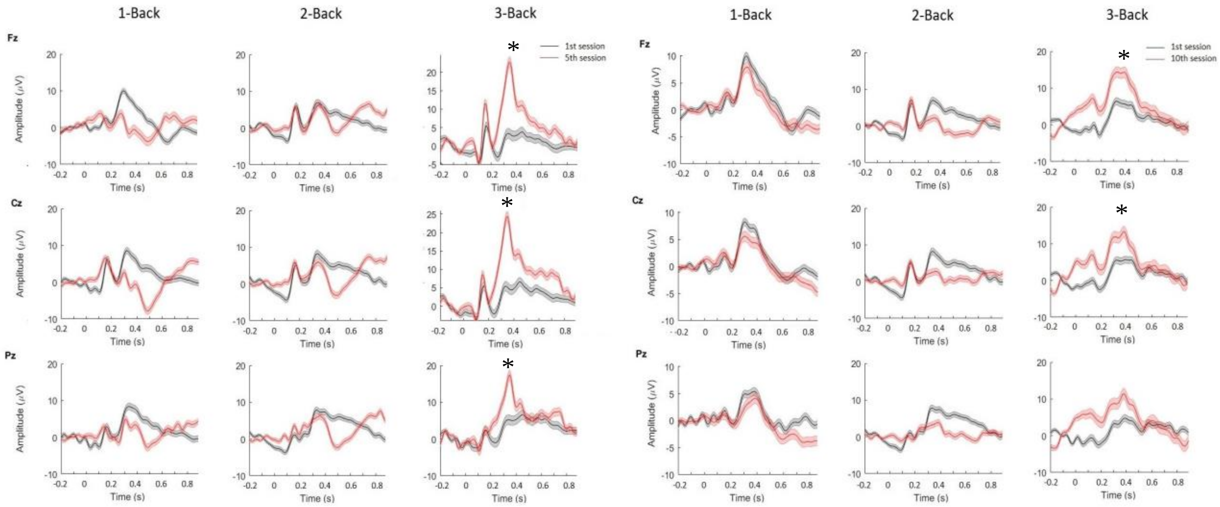


Figure 6. P300-ERPs amplitude (target minus non-target) in 9 subjects of the CTG-young group. Differences between 1st and 5th session (middle session, left), and between 1st and last session (10th session, right) are shown for channels Fz, Cz, and Pz (row-wise). Significance was measured using three-way ANOVA ($p < 0.05$). Error bars indicate SEM.

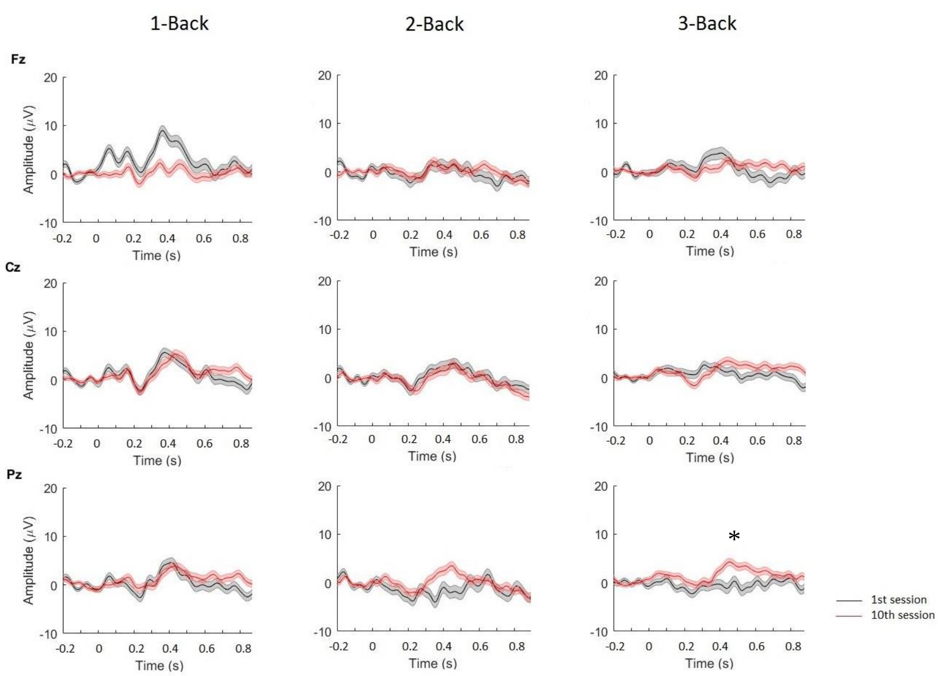


Figure 7. P300-ERPs amplitude in 10 subjects of the CTG-older group. Differences between 1st and last session (10th session) are shown for channels Fz, Cz, and Pz (row-wise). Significance was measured using three-way ANOVA ($p < 0.05$). Error bars indicate SEM.

For healthy older subjects, we found significant effects for accuracy in N-Back task between CTG and PCG ($F_{(1)}$=7.26, p<0.05) for group, and in TOVA between CTG and PCG ($F_{(1)}$=30.88, p<0.001) for group. No significant differences in CORSI and RAVEN accuracies were found between groups. Significant effects were found for RT between CTG and PCG for the N-Back task, for training x N-Back level interaction ($F_{(2)}$=3.54, p<0.05).

*

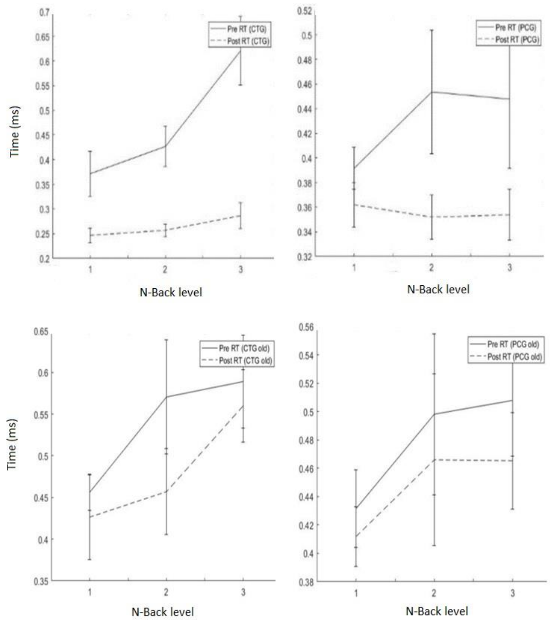


Figure 8. RT (correct responses) in the performance from pre to post-test in the *n*-back task of 2 groups of young (top row) and older subjects (bottom row): CTG (left column), and PCG (right column). Error bars indicate SEM.

Since Dahlin et al. [34] showed that younger adults gain more benefits from cognitive training than older adults, but Bherer, Westerberg, and Bäckman [36] showed the opposite (older subjects gained more positive effects than younger ones), we analyzed the differences between young and older adults. We used a multivariate ANOVA (MANOVA) for the factors group (young and older) in the CTG, and session (pretest and post-test, Figures 10 and 11). Significant effects for accuracy in CORSI test ($F_{(1)}=6.18$, $p<0.05$) for group, as well for RAVEN ($F_{(1)}=24.97$, $p<0.001$) for group, and in training ($F_{(1)}=5.7$, $p<0.05$). No significant differences in the N-Back and TOVA results between groups were found. For the N-Back task (Figure 13), significant effects were found for RT between CTG young and older for group ($F_{(1)}= 27.98$, $p<0.001$), for training ($F_{(1)}=24.83$, $p<0.001$), for N-Back task ($F_{(2)}=9.12$, $p<0.001$), and for age (group) x training effects interaction ($F(1)=8.06$, $p<0.05$).

TABLE I. PERCENT CORRECT (MEANS AND STANDARD DEVIATION) OF PRE- AND POST-TEST PERFORMANCE (ACCURACY) BETWEEN TRAINING GROUP AND PASSIVE CONTROL GROUPS IN TRAINED (N-BACK) AND UNTRAINED TASKS, FOR HEALTHY YOUNG AND OLDER SUBJECTS.

	Young subjects				Older subjects			
Task	Cognitive training group		Passive control group		Cognitive training group		Passive control group	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
N-Back	92.76±3.36	98.33±1.46	94.66±4.36	95.66±2.5	94.66±1.66	97.66±0.86	92.9±2.06	92.33±3.6
TOVA	84.4±10.4	93.2±4.2	85.2±8.8	90±6.4	92.6±3	96±3.4	106±5.2	106.4±3.4
CORSI	58.66±6.66	69.33±16.66	61.33±14.66	62.66±18.66	46.66±6.66	55.33±7.33	58±11.33	60±7.33
RAVEN	94.16±6	97±4.6	92.33±8.66	93±6.66	69.3±15.5	84.33±1.83	82.83±8	85±5.83

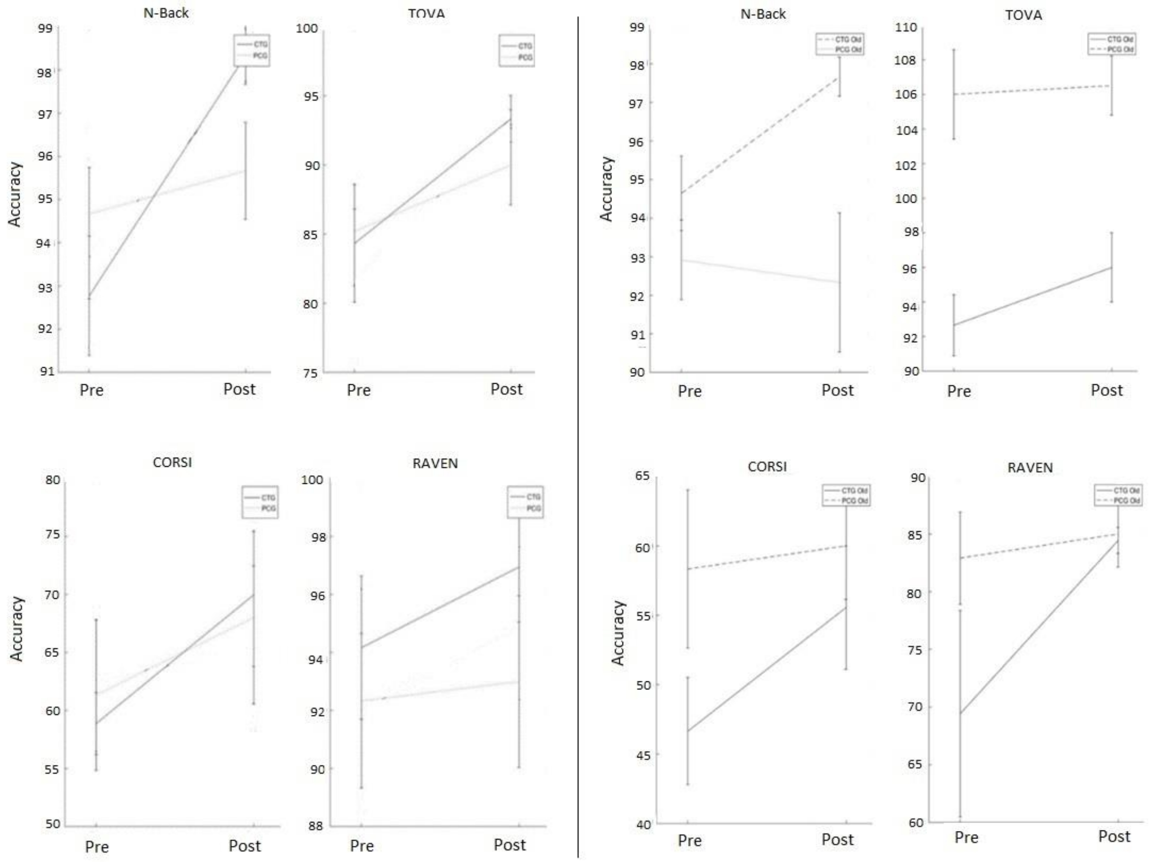


Figure 9. Left: pre- to post-test correct performance (in %) of CTG and PCG groups of healthy young subjects for the N-back task, TOVA test, CORSI test, and RAVEN test. Right: idem but for the 2 groups of healthy older subjects. Error bars indicate SEM.

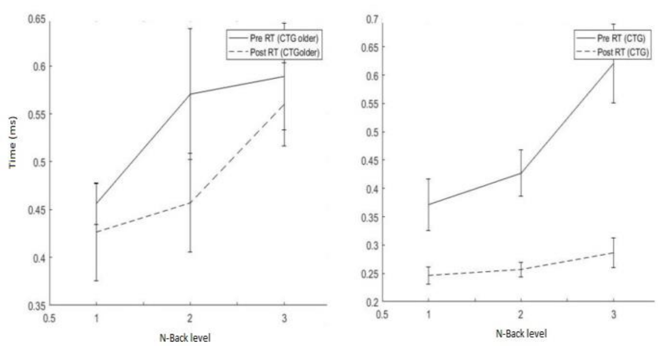


Figure 10. RT for correct responses in pre- and post-test of the N-back task between CTG groups of young (left panel) and older subjects (right panel). Error bars indicate SEM.

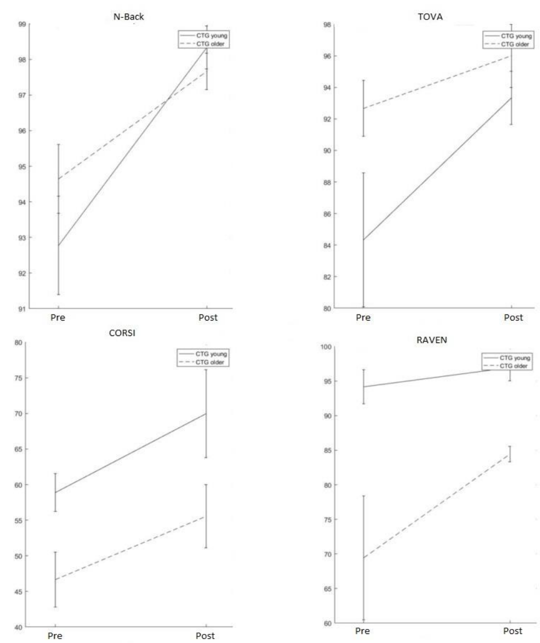


Figure 11. Top row: percent correct performance of CTG young and older) for pre- to post-test of the N-Back task (left) and TOVA test (right). Bottom row: idem but for the CORSI- (left) and RAVEN test (right). Error bars indicate SEM.

In summary, with our pre- and post-training tests on other untrained cognitive functions, we wanted to verify whether any transfer effects could be obtained after N-Back training. Our results indicate clear improvements in attention, which is in line with the outcomes of Dahlin et al. [34] who showed larger improvements for younger adults than for older ones.

IV. DISCUSSION

We investigated whether cognitive training using an N-Back task only improves performance in this task or transfers to other tasks. To assess this, we performed 10 N-Back training sessions in one group (CTG) of young and older adults, and assessed their cognitive performance with a battery of cognitive tasks (TOVA, CORSI and RAVEN test) before and after training. During training, CTG participants performed the 1,2,3-Back task. Furthermore, a second group of participants (PCG) for young and older adults performed no training but was subjected to the same battery of cognitive tests. We found for healthy young subjects that training indeed improves performance of the CTG group compared to the PCG group. Therefore, there is a clear improvement for CTG on the task they were trained on (N-Back task). In contrast, the transfer of training effects into other tasks is more nuanced: the transfer was significant for the attention (TOVA) test, but only for CTG, but there was a trend for spatial memory (CORSI) and fluid intelligence (RAVEN) tests that was larger for CTG than for PCG. These results are in line with the conclusions of Jaeggi et al. [14] who showed that, when using a working memory task, not only working memory improves but also fluid intelligence. However, the results are not in line with those of Dahlin et al. [34] who found that, whereas working memory training improves performance of a related working memory task, other cognitive functions did not improve.

For healthy older adults, we found a significant improvement in N-Back task performance for CTG compared to PCG, and in TOVA task performance for CTG compared to PCG, and a trend of improvement in CTG for the CORSI and RAVEN tests compared to PCG. These results are in line with the studies of Yang et al. [37] and Dahlin et al. [34] who showed that the capacity of plasticity in the ageing brain improves cognitive functioning. One explanation for transfer effects being significant only for the attention task (TOVA), and not for the spatial memory (CORSI) and fluid intelligence tasks (RAVEN), for both groups (young and older), maybe due to our small sample size.

Furthermore, we wanted to compare the effect of cognitive training in young and older subjects in CTG and PCG, and P300 ERP amplitude during N-Back task performance in CTG, for both groups. The results of CTG showed that cognitive training improves performance of the N-Back task in RT, and that the transfer effects were significant for the RAVEN and CORSI tests, but not for TOVA, which is in line with the Jaeggi et al. [14] results. Although both young and older healthy subjects improved their performance after WM training, younger healthy subjects showed a larger effect following cognitive training compared to older healthy subjects, which is in line with the results of Brehmer et al. [20] and Dahlin et al. [34]. We observed significant differences in P300 amplitude between young and older subjects for N-Back level of the task, target stimuli and sessions (first vs last, first vs middle) in 3-Back task, indicating that 5 training sessions could be enough to improve the trained WM task. In this way, we complemented the study of Friedman et al. [33] who used a simple oddball paradigm to observe the differences of ERP amplitude in young and older adults, showing the number of training sessions that could be necessary to improve the N-Back task.

An issue that deserves consideration is why N-Back training in our study did not produce transfer effects in CORSI test (spatial memory), while in Dahlin et al. [34] study they observed transfer effects to another memory task (near-transfer effects). Furthermore, as the EEG results from our study suggest a change in the ERPs-P300 during cognitive training, our future study will consider not only behavioral data (accuracy and RT) but also P300-ERPs to change in real time the difficulty level of the task, in order to avoid fatigue or boredom for the subject.

V. CONCLUSION

As cognitive training is becoming important during one's life-span, to maintain cognitive plasticity and postpone cognitive decline, we decided to study the power of this tool in search of the benefits that can be gained by young and older adults. We showed that N-Back training not only improves WM but also transfers to another cognitive function (i.e., attention). The results provide evidence that it is indeed possible to improve the trained task (working memory), but also transfer to other cognitive tasks. Furthermore, our findings are promising to be used with a larger cohort of healthy older subjects and patients with cognitive decline.

ACKNOWLEDGMENT

VP is supported by research grant No. G088314N, sponsored by the Belgian Fund for Scientific Research-Flanders, MMVH by research grants No. PFV/10/008 from the Program Financing program, No. IDO/12/007 from the Interdisciplinary research project of KU Leuven, No. IOF/HB/12/021 sponsored by the Industrial research fund project *of* KU Leuven, No. G088314N and No. G0A0914N sponsored by the Belgian Fund for Scientific Research-Flanders, No. IUAP P7/11 from the Interuniversity Attraction Poles Programme – Belgian Science Policy, No. GOA 10/019 from the Flemish Regional Ministry of Education (Belgium) and No. AKUL 043 sponsored by the Hercules Foundation. BW is supported by the Flemish Agency for Innovation and Entrepreneurship (VLAIO).

REFERENCES

[1] V. Pergher, B. Wittevrongel, J. Tournoy, B. Schoenmakers, J. Arsenault, and M. M. Van Hulle "N-Back training and transfer effects in healthy young subjects using EEG", BRAININFO 2017, The Second

International Conference on Neuroscience and Cognitive Brain Information, IARIA, July 2017, pp. 1-7, ISSN: 2519-8653, ISBN: 978-1-61208-579-1.

[2] S. C. Li, F. Schmiedek, O. Huxhold, C. Röcke, J. Smith, and U. Lindenberger, “Working memory plasticity in old age: practice gain, transfer, and maintenance”, Psychol. Aging 23, pp. 731–742, 2008.

[3] A. Baddeley, “Working Memory and Conscious Awareness”, Theories of memory, pp. 11-20, 1992.

[4] J. H. Smith-Spark, and J. E. Fisk, “Working memory functioning in developmental dyslexia”, Memory, 15(1), pp. 34-56, 2007.

[5] T. Klingberg, “Training and plasticity of working memory”, Trends Cogn Sci 14, pp. 317–324, 2010.

[6] C. C. von Bastian, and K. Oberauer, “Effects and mechanisms of working memory training: a review”, Psychol Res 78, pp. 803–820, 2014.

[7] J. Au, E. Sheehan, N. Tsai, G. J. Duncan, M. Buschkuehl, and S. M. Jaeggi, “Improving fluid intelligence with training on working memory: a metaanalysis”, Psychon Bull Rev 22, pp. 366 –377, 2015.

[8] W. K. Kirchner, “Age differences in short-term retention of rapidly changing information”, J Exp Psych, pp. 1-17, 1958.

[9] J. F. Mackworth, “Paced memorizing in a continuous task”, Journal of Experimental Psychology, 58(3), p. 206, 1959.

[10] A. S. Gevins, S. L. Bressler, B. A. Cutillo, J. Illes, J. C. Miller, J. Stern, and H. R. Jex, “Effects of prolonged mental work on functional brain topography”, Electroencephalography and clinical neurophysiology, 76(4), pp. 339-350, 1990.

[11] M. J. Kane, and R. W. Engle, “The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective”, Psychonomic bulletin and review, 9(4), pp. 637-671, 2002.

[12] A. M. Owen, K. M. McMillan, A. R. Laird, and E. Bullmore, “N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies”, Human brain mapping, 25(1), pp. 46-59, 2005.

[13] J. A. Schneiders, B. Opitz, C. M. Krick, and A. Mecklinger, “Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation”, Cerebral Cortex, 21(11), pp. 2555-64, 2011.

[14] S. M. Jaeggi, “Improving fluid intelligence with training on working memory”, Proc. Natl. Acad. Sci. 105, pp. 6829– 6834, 2008.

[15] S. M. Jaeggi, “The relationship between n-back performance and matrix reasoning implications for training and transfer”, Intelligence 38, pp. 625–635, 2010.

[16] E. Dahlin, L. Nyberg, L. Bäckman, and A. S. Neely, “Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance”, Psychol. Aging 23, pp. 720–730, 2008.

[17] P. B. Baltes, U. Lindenberger, and U. M. Staudinger, “Life span theory in developmental psychology”, Handbook of child psychology: Theoretical models of human development (Vol. 1, pp. 569-664). Hoboken; 2006.

[18] E. Bialystok, and F. I. Craik, (Eds.) “*Lifespan cognition: Mechanisms of change*”, Oxford University Press, USA; 2006.

[19] A. B. Morrison, and J. M. Chein, “Does working memory training work? The promise and challenges of enhancing cognition by training working memory”, Psychonomic Bulletin & Review, 18, pp. 46–60. 2011.

[20] L. Bherer, A. F. Kramer, M. S. Peterson, S. Colcombe, K. Erickson, and E. Becic, “Training effects on dual-task performance: Are there age-related differences in plasticity of attentional control?”, Psychology and Aging, 20 (4), pp. 695–709, 2005.

[21] J. Chein, and A. Morrison, “Expanding the mind's workspace: Training and transfer effects with a complex working memory span task”, Psychonomic Bulletin and Review, 17 (2), pp. 93–199, 2010.

[22] E. Dahlin, L. Backman, A. S. Neely, and L. Nyberg, “Training of the executive component of working memory: Subcortical areas mediate transfer effects”, Restorative Neurology and Neuroscience, 27 (5), pp.405–419, 2009.

[23] Y. Brehmer, H. Westerberg, and L. Bäckman, “Working-memory training in younger and older adults: training gains, transfer, and maintenance”, Training-induced cognitive and neural plasticity, p. 72, 2012.

[24] L. L. Richmond, A. B. Morrison, J. M. Chein, and I. R. Olson, “Working memory training and transfer in older adults”. Psychology and aging, 26(4), p. 813, 2011.

[25] L. M. Greenberg, and I. D. Waldmant. "Developmental normative data on the test of variables of attention (TOVA™)." Journal of Child Psychology and Psychiatry 34 (6), pp. 1019-1030, 1993.

[26] R. P. Kessels, M. J. Van Zandvoort, A. Postma, L. J. Kappelle, and E. H. De Haan, “The Corsi block-tapping task: standardization and normative data”, Applied neuropsychology, 7(4), pp. 252-258, 2000.

[27] J. Persson, and P. A. Reuter-Lorenz, “Gaining Control: Training Executive Function and Far Transfer of the Ability to Resolve Interference [retracted]”, Psychological Science, 19(9), pp. 881-888, 2008.

[28] X. Zhao, Y. Wang, D. Liu, and R. Zhou, “Effect of updating training on fluid intelligence in children”, Chinese Science Bulletin, 56(21), pp. 2202-2205, 2011.

[29] J. Karbach, and J. Kray, “How useful is executive control training? Age differences in near and far transfer of task-switching training’, Dev Sci 12, pp. 978–990, 2009.

[30] R. J. Croft, and R. J. Barry, “Removal of ocular artifact from the EEG: a review”, Neurophysiologie Clinique/Clinical Neurophysiology, 30(1), pp. 5-19, 2000.

[31] S. Watter, G. M Geffen, and L. B. Geffen, “The n-back as a dual-task: P300 morphology under divided attention”, Psychophysiology, 38(06), pp. 998-1003, 2001.

[32] A. M. Brouwer, M. A. Hogervorst, J. B. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, “Estimating workload using EEG spectral power and ERPs in the n-back task”, Journal of neural engineering, 9(4), p. 045008, 2012.

[33] D. Friedman, G. V. Simpson, “ERP amplitude and scalp distribution to target and novel events: effects of temporal-order in young, middle-aged and older adults”, Cogn Brain Res. 2: pp. 49–63, 1994.

[34] E. Dahlin, A. S. Neely, A. Larsson, L. Backman, and L. Nyberg, “Transfer of learning after updating training mediated by the striatum”, Science 320, pp. 1510–1512, 2008.

[35] A. Gevins, and M. E. Smith, “Neuropsychological measures of cognitive workload during human-computer interaction”, Theoretical Issues in Ergonomics Science, 4(1-2), pp. 113-131, 2003.

[36] T. Salminen, P. Frensch, T. Strobach, and T. Schubert, “Age-specific differences of dual n-back training”, Aging, Neuropsychology, and Cognition, 23(1), pp. 18-39, 2016.

[37] L. Yang, R. T. Krampe, and P. B. Baltes, “Basic forms of cognitive plasticity extended into the oldest-old: retest learning, age, and cognitive functioning”, Psychology and aging, 21(2), p. 372, 2006.

A Complete Set-up to Evaluate the Correlation Between Blood Pressure and Pulse Transit Time

Adhurim Hajzeraj, Marco Belcastro, Davide Alfieri, Brendan O'Flynn
Tyndall National Institute
University College Cork
Cork, Ireland
E-mail: adhurim.hajzeraj@tyndall.ie; marco.belcastro@tyndall.ie; davide.alfieri@tyndall.ie; brendan.oflynn@tyndall.ie

***Abstract*— Blood pressure (BP) has always been one of the most important parameters in monitoring cardiovascular system conditions and coronary artery diseases (CAD), such as angina and myocardial infarction (commonly known as a heart attack). This is due to the fact that many of the changes within the cardiovascular system, such as clogged arteries, for example, are reflected by changes in BP. A number of methods and devices that can measure BP are available on the market for both clinical and consumer use. However, being able to measure one's own BP non-invasively, with the required frequency (even continuously) in a comfortable fashion remains an unsolved problem using currently available systems. To date, the Pulse Transit Time (PTT) measurement method has been seen as a feasible approach to help bring current blood pressure monitoring systems to a stage where non-invasive, continuous measurements are viable. However, developing a system which uses the PTT method for blood pressure measurement is as yet an unsolved problem and it remains a challenge to achieve accurate BP results despite considerable research in the past decade. In this paper, we present the first step in building a smart sensing system that overcomes the technical difficulties associated with accurate measurement of PTT. The novel hardware developed incorporates multi-modal sensing capability to explore and quantify the relationship between blood pressure and PTT. The evaluation system is completed by efficient, simple and fast embedded software algorithms, user interface, and clinical validation trials that will enable delivering a novel PTT-based blood pressure monitor.**

Keywords - blood pressure; pulse transit time; ECG; PPG; calibration; real time data; clinical trials.

I. INTRODUCTION

According to statistics, cardiovascular diseases (CVD) are the main cause of deaths in Europe with 45% of all deaths caused by CVDs [1]. The overall estimated cost to the EU economy is €210 billion a year [2]. The motivation for this research is to reduce these statistics by finding an improved method of continuously monitoring real time blood pressure (BP). This will help clinicians to monitor, diagnose and improve the condition of the cardiovascular system [3] through the availability of more detailed real time data sets.

The current state of the art in BP measurement utilises a number of different methods and devices including catheterization, auscultation, oscillometry, volume clamping, and tonometry, with catheterization being the most accurate standard currently being used [4]. In general, the accuracy of the measurements obtained by existing devices is acceptable, but they have a number of drawbacks. Firstly, where inflatable cuffs are used, they tend not to be comfortable for the user and as a result are inappropriate for long term continuous monitoring. Secondly, clinical grade systems need to be operated by doctors, which causes a phenomenon called 'white coat syndrome', where BP readings are inaccurate due to the presence of the clinicians.

A system which would enable clinicians to take accurate, real time and continuous BP measurements would be invaluable to doctors in diagnosing CVDs at an early stage [3].

In this paper, the first development stage of such a system for BP monitoring based on the PTT method is presented. Section II of this paper describes the principles behind the PTT measurement method. Section III describes the evaluation of components integrated into the system hardware platform and the design of a custom hardware and new sensors.

Sections IV and V describe the developed embedded software and user interface respectively. The experimental protocol which was developed in conjunction with clinical inputs and which was used in the validation trials and the associated measurements taken from the subjects as well as the initial processed results are shown in Section VI. Section VII presents the second design phase of the final device which is currently underway. Finally, Section VIII concludes the work with some preliminary test results and a description of future work to be undertaken.

II. PULSE TRANSIT TIME

A method that proposes to have a good potential to enable non-invasive continuous BP measurement is the Pulse Transit Time method. PTT is defined as the time needed for the blood wave that goes out of the heart with each beat to arrive at a peripheral body site, in our case the wrist or fingertip. The delay (PTT) is calculated as the time difference between the peak of electrocardiograph (ECG) and the peak of photoplethysmogram (PPG) signals.

The main factors that determine the speed of propagation of the pulse wave, and which thus affect the PTT value, are the elasticity coefficient, the thickness of the arterial wall, the end-diastolic diameter of the vessel lumen and blood density [5]. In 1878, Moens and Korteweg developed a formula that relates the velocity of the pulse with the factors as described in (1):

$$PWV = \frac{D}{PTT} = \sqrt{\frac{tE}{\rho d}} \quad (1)$$

The Pulse Wave Velocity (PWV) is dependent on a number of arterial properties, namely the elasticity of the artery wall E, the arterial wall thickness t, the arterial diameter d and the density of the blood ρ [6]. So, as the density of blood is close to that of the density of the water, the main factors that influence the velocity of the pulse wave are the properties of the arterial vessels, stiffness and thickness. These factors vary from person to person on an individual basis [5]. In the calculation of BP parameters from PTT readings for an individual, this variation would be addressed through a calibration activity as part of the measurement protocol, but this is insufficient, as vessel properties keep changing in a dynamic fashion due to a variety of factors [7]. Factors that can change vessel properties include ambient temperature [8, 9] barometric pressure [10], sleep-wakefulness status of the person [11], time of the day [12], sport activities [13] and sometimes it can even change by the control of the brain or sympathetic system [14].

III. TYNDALL PTT EVALUATION SYSTEM

In order to evaluate the impact of the variability inducing factors described above in the evaluation of BP and blood vessel properties, the WSN group at Tyndall has developed novel multi modal sensing system with the required hardware, sensors and algorithms, to be able to carry out the necessary measurements. The first implementation of the measurement system is shown in Figure 1.

A. Microcontroller Hardware System Design

The focus of this system implementation is the development and test of a smart sensing device in the form factor of a wrist watch, which would include the processor, battery, data visualisation interface, communications and all sensors required to measure BP using the PTT method.

To achieve this, our design methodology had to focus on three main parameters. Firstly, the size of the components should be small enough to fit in our miniaturised target device. Secondly, the microcontroller (MCU) should be powerful in its operations, as it will be a real time data acquisition system and it will run all the algorithms inside the embedded microprocessor. And thirdly, it should be a system that spends as little energy as possible as it is battery operated.

Considering all these parameters, we have chosen the STM L series microcontrollers [15] to use as a computing unit.

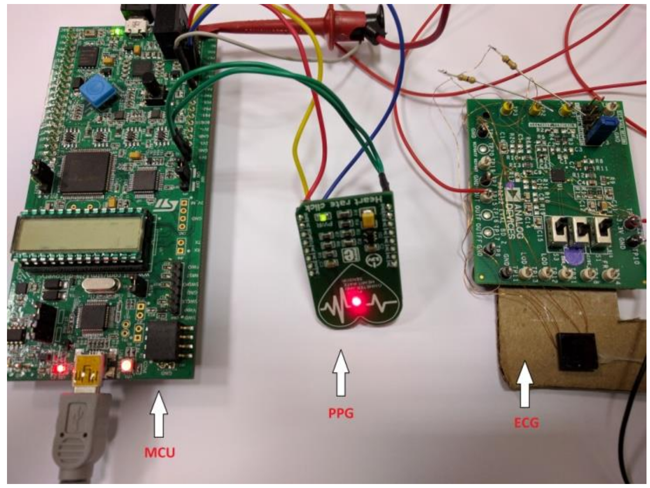


Figure 1. Setup of the evaluation board prototype data acquisition system.

As this device is a 32-bit microprocessor that can run at clock speeds of up to 100 MHz, it will satisfy our computing needs from an algorithmic perspective. At the same time, size-wise the STM component is small enough to fit inside a smart watch form factor system and is one of the lowest energy consuming microcontrollers on the market, if embedded code is managed appropriately. In addition to the before mentioned performance characteristics of the STM component, the MCU is also required to have a powerful Analog to Digital converter (ADC) and other digital interfaces to read data from the chosen sensors. Figure 1 shows a picture of evaluation boards that we are using as a first data acquisition prototype. The board on the left is the evaluation board with the STM microcontroller MCU.

B. ECG and PPG sensors

To develop the necessary algorithms to calculate PTT, it is anticipated that we will need to have datasets associated with two signals associated with the cardiovascular system. The heart, during work activities, generates a bio-signal that is well known and characterised as ECG, this is one of the waveforms we need to establish our BP measurement algorithm. Generally, these signals are recorded by placing electrodes on the skin, on the chest area, where the signals are stronger. Reading ECG by placing electrodes on the chest, gives the strongest and most easily read signal, but in general, this is not comfortable and convenient for the user. For the Tyndall ECG measurement system, we will use two active electrodes, which are placed in the watch in a wrist mounted implementation.

The sensor used for this application is named the Electric Potential Integrated Circuit (EPIC) from Plessey Semiconductors [16]. The EPIC sensor can be used, as a replacement technology for traditional wet-electrode ECG pads, because it requires neither gels nor other contact-enhancing substances. When the EPIC sensor is placed on (or in close proximity to) the patient, an ECG signal can be recovered.

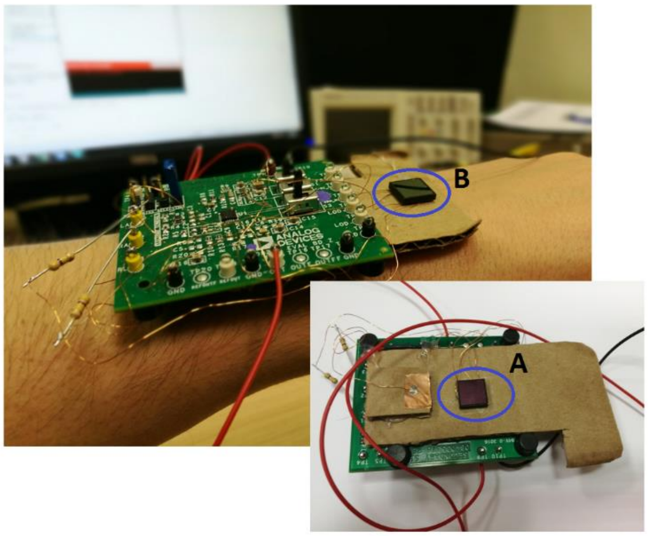


Figure 2. Evaluation board of ECG sensor.

To illustrate the placement of the electrodes to enable ECG measurements, consider Figure 2, where we show the development board of ECG sensor placed on a subject's wrist (electrodes in contact with the skin on the underside of the board as shown in the smaller picture). To enable ECG measurement in the wrist mounted scenario, there are two electrodes, electrode A and B. Electrode A is placed under the watch and will touch the skin. Electrode B is placed over the watch, so every time the user wants to measure BP, the user should touch electrode B with a finger of the opposite hand to read the differential signal. Figure 3 shows the differential amplifier which enables generation of the differential bio-signal. Input1 is the electrode touching the skin on the wrist and Input2 is the electrode touched by the finger of the opposite arm.

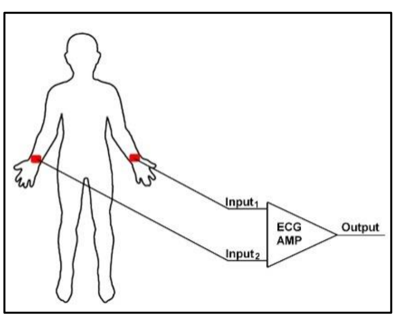


Figure 3. Generating differential biosignal from two inputs

The second waveform that is needed to develop the BP measurement algorithm is the signal generated from a photoplethysmogram (PPG) sensor, which shows the level of the volume of blood circulating near the sensor. The system used in the Tyndall implementation is a Maxim PPG sensor, MAX30100 [reference], which contains two light-emitting diodes (LED), one red and the other one is infrared together with a photodiode.

The change in volume caused by the pressure pulse is detected by illuminating the skin with the light from a LED and then measuring the amount of light either transmitted or reflected to a photodiode. Each cardiac cycle appears as a peak, as seen in Figure 7.

We are using this sensor to read PPG data from the fingertip, which is optimal as at the fingertip the PPG signal tends to be clear and not very noisy. Care needs to be taken however to ensure the sensor does not move when it is touching the fingertip. For initial data set acquisition and algorithm development, the fingertip implementation will be used for this reason until the wrist PPG sensor development is finished.

The same sensor has been tested acquiring real time data from the wrist mounted implementation. On the wrist the waveforms tend to be noisier and will require further filtering and signal processing to develop a signal of sufficient quality for use in real scenarios.

C. *Prototype wrist mounted PPG data acquisition system*

Based on the PPG measurement system described in the previous section, the WSN group at Tyndall have developed an application specific, new, PPG sensor system, which will enable data sets from the wrist to be taken directly and is of a form factor appropriate to that envisaged for the final product. A picture of the board is shown in Figure 4, where one can see the main components on a 5cm by 3cm sized PCB microsystem. The board contains a USB connection module, the microcontroller, the ECG sensor and the new PPG sensor circuit, which are described in the next few paragraphs. A 3D printed enclosure has been printed for the board, which makes it able to be used as a wrist mounted system and facilitate collection of data from participants in a reliable, repeatable and accurate fashion.

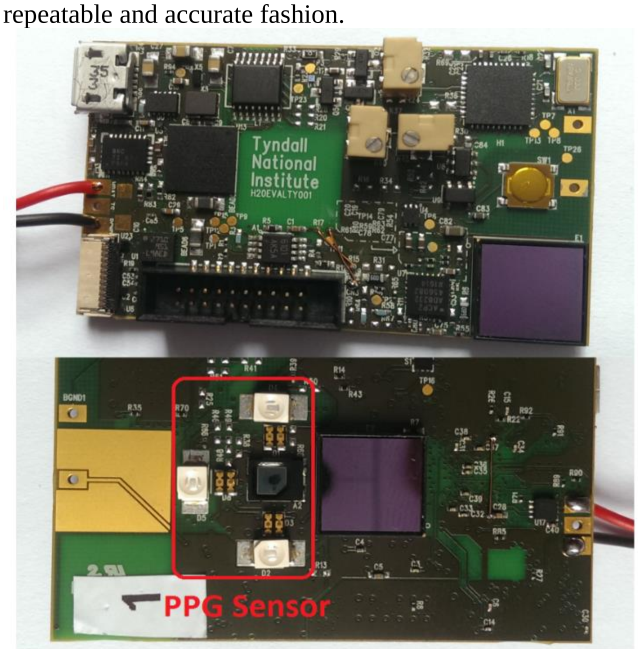


Figure 4. Wrist mounted PPG data acquisition system.

Experiments have shown that variation in the position of the LEDs, as well as the particular power and wavelengths of different LEDs impact significantly on the quality of the PPG signal obtained.

To evaluate different placement scenarios, we have designed an integrated circuit with nine LEDs mounted in a circular manner around a photo diode, three green LEDs, three red and three infrared are used for this experimental setup. Diodes are arranged around the photodiode, as shown in Figure 4. This series of experiments is currently underway and will enable the design team to find the optimal configuration of diodes from the perspective of position, colour (wavelength) and intensity.

The new PPG sensor circuit is designed with the intention to test different configurations, communication protocols and positions of LEDs and photodiode within the same board. There are three experimental setups implemented in the same board for this system to optimise PPG acquisition parameters. A block diagram of this circuit is shown in Figure 5, which describes the flow, how the MCU can control LEDs through the three LED controlling plans.

In option A (Plan A), an additional integrated circuit (IC) to MCU will control which LEDs will be on and the limit of current intensity. This IC has an analog output so that we can vary very precisely the intensity of the light output from the LEDs to achieve the optimum level for wrist monitoring of PPG signals.

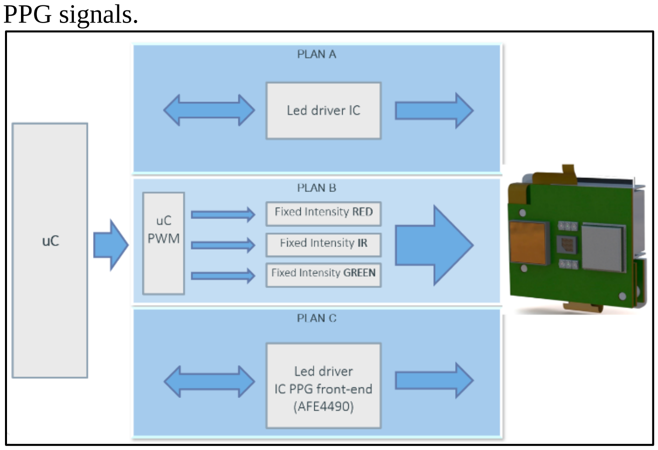


Figure 5. Three different test configurations of LEDs from PPG sensor.

In the option B (Plan B), every colour group (green, red and infrared) is connected to a potentiometer and Pulse Width Modulation (PWM) unit of the MCU. This will enable variation of the intensity of the LEDs very precisely. Also in this case, the output is analog and the ADC can be used to read data.

The option C (Plan C) requires the integration of an additional IC also. This chip (Analog Front End - Texas Instruments AFE4490 [17]) controls the state of LEDs in a similar way to option A, except that LEDs here can be controlled only in groups, so there is the same intensity for all the red LEDs for example. In this case, the output is digital, and eliminates the usage of the ADC.

D. Sensor Data Acquisition System evaluation

To calculate PTT, we use the peak of the ECG waveform and the peak of the PPG wave. Other features of waves can be used, like the beginning of QRS complex in ECG or the segment with the highest slope in PPG signal, but peaks are easier to detect and provide the fastest route to an initial prototype implementation for clinical tests. Additional tests in the future can carry on to determine if other features of the waves may give similar or improved accuracy. To identify the required peaks in the signals an initial sampling frequency of 100Hz was used. The sampling rate can be increased in future experiments if additional features need to be detected which will require higher precision [18]. The Maxim PPG sensor has a digital output using an SPI protocol and it generates an interrupt after each sample is taken. The USB virtual COM port is enabled in the microcontroller, so we use this interface to send data to and from a PC to develop the data analytics. The PPG and ECG data sets are sent in real time to the PC, and plotted to check the quality of the waves. As is described below in this section, the waveforms from the sensors when located on a fingertip are of superior quality and are what we are using for the initial algorithm development in the calculation of PTT.

We tested the same sensor on the wrist. It was quickly apparent that the sensor needs much more time for the data readings from the wrist to stabilize as compared with when it is worn on fingertip. Even after we start to see the standard shape of the wave, it is very noisy and hardly readable, and also we lose the waveform with small movements of the sensor or hand. In Figure 6, we can see the best wave we could get from the Maxim sensor when worn on the wrist. For the final wrist mounted system, additional software filtering and signal processing are required on the waveforms for this reason.

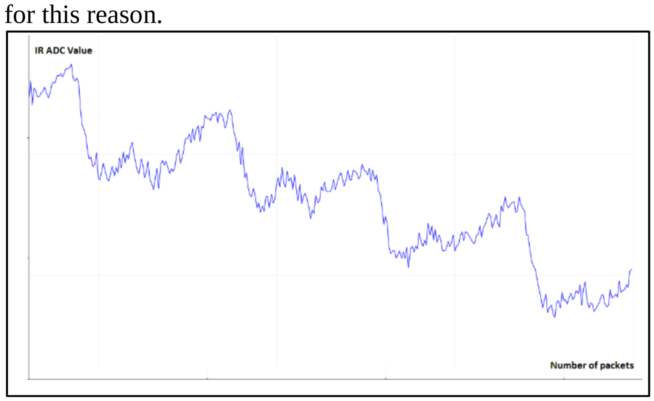


Figure 6. Real time PPG data from Maxim sensor on wrist.

The ECG sensor is combined with a circuit created using an Analog Devices (AD8232 [19]), which was designed to record ECG signals using classic electrodes, but is modified and used with EPIC ECG electrodes in our case. This chip has an analog output and is sampled at the same rate, 100Hz, as the PPG wave. The MCU's ADC unit is used, and the data from ADC registers are read every time an interrupt is triggered by the PPG sensor. The same sampling rate (of 100

Hz) for both waves is used in the literature [20], and we have also implemented this, which is sufficient to evaluate the system during this initial phase. It means that we will use only one interrupt service routine in the microcontroller, and less instructions will be executed to run the algorithms and measure time difference between peaks of two waves.

As seen in Figure 7, the red colour wave is the real time ECG signal from the sensors. The signal is of sufficient quality to be able to detect the peaks as this is the main feature we will extract from ECG at this stage. This signal is plotted from raw data, so no post processing has been implemented. In next revision of the software, software noise filtering can be applied on the signal, which would result with a clearer graph, higher accuracy and other features of the signal can be detected. In Figure 7 again, we can see the plotted waveforms and datasets needed to determine PTT. The blue, lower graph is the PPG waveform from the fingertip. This wave from fingertip is clear and consistent, which means peaks can be easily located. Both waves are shown on the same plotting area to visualize PTT. Time difference between two vertical segments is the value of PTT.

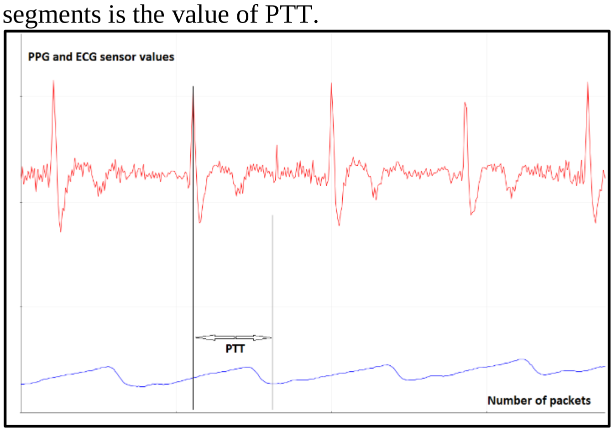


Figure 7. Real time ECG and PPG data on the same plot and PTT.

Figure 8 shows a close up representation of the ECG peaks QRS complex. The R peak is considered a starting point to measure PTT for the purposes of our algorithm development. The PPG wave is inverted, as the wave shows the level of absorption, so it means at the lowest peak the light is absorbed the most and this is because of the volume of the blood in vessels at that moment.

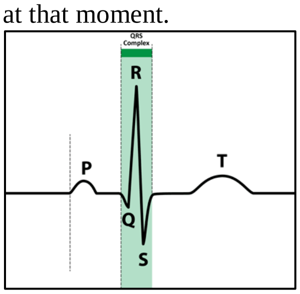


Figure 8. ECG QRS complex.

So, the distance between the two black vertical lines on Figure 7 is the value of PTT used in our evaluation and future calculations of blood pressure changes.

IV. FIRMWARE AND CALCULATION OF PTT

In the previous section were shown results from the testing of the hardware and software of the initial evaluation system. With the exception of the PPG sensor, the design of the Tyndall system designed as a custom PCB board (H2O device) the same hardware components as described in Section III are used. The PPG sensor designed by the Tyndall team and described in Section III is used in the H2O device in place of the Maxim PPG sensor. Using the new PPG sensor designed we were able to record less noisy PPG signal waveforms from the wrist. The improvement in signal to noise ratio can be seen if the PPG signal shown in Figure 6 is compared with the PPG data shown in Figure 10, where the PPG signal as measured on the wrist by the Maxim sensor and the Tyndall sensor are shown respectively. The new PPG sensor is also used while performing clinical measurements on 33 participants as part of the system validation trials, and in all cases, there were no issues on reading sufficiently clear PPG signal. Results of calculation of PTT during clinical trials are shown in Section VI, Table I.

In the first version of the software, PPG and ECG are recorded at 1 kHz sampling rate compared to 100Hz in the previous version. A sampling rate of 100Hz has a 10ms time difference between two points, which lower the accuracy of the estimation to approximately ±2mmHg.

Sampling time is measured by timer counter interrupts. The counter interrupt is also used to measure the time between samples and to convert the number of samples to values of milliseconds. Windows of 4 seconds are used to collect the data, which afterwards are processed while the next window of data is collected. We have used windows of duration of 4 seconds two reasons. Firstly, in this time period it is generally possible to locate at least three heartbeats in one window, enabling the calculation of heart rate and PTT. Secondly, a longer window would delay the process of calculating PTT. When the user touches the ECG electrode, the sensor requires some time to start reading the ECG waveform. As averaging is used to localize the R complex, most of the time algorithms would be unable to calculate PTT. This mean that calculation will be performed in the next window. If the window is longer, overall execution would take more time to finish. Collecting samples and running the algorithms simultaneously means that in between sample readings, the processing power is used to execute parts of the algorithm with the data from the previous window. In parallel, after completion of one window, collected and processed data are transmitted through serial COM port. Transmitted data is formatted so as to conform to the communication protocol, designed to communicate to the interface, which is described in the following section.

To derive systolic and diastolic BP from PTT it is required to locate three points on ECG and PPG, therefore to determine systolic PTT (SPTT) and diastolic PTT (DPTT) [21, 22]. The time difference between the peak of QRS complex of ECG, labelled with 'A' in Figure 9, and the minima of PPG signal, labelled with 'C' is marked as SPTT.

And the difference between the peak of QRS complex and the maxima of the PPG signal, labelled with ‘B’ in Figure 9, is labelled as DPTT.

The peak of QRS is located by averaging the ECG signal three times and setting the value of averaging as a threshold. Points ‘B’ and ‘C’ are located by exploring PPG signal in a time distance range where PTT value usually is. The algorithm will start scanning for minima first. Once it establishes the minima, the algorithm will continue to search for the point of the inflexion on PPG signal, which should be between minima of PPG and ECG peak. Knowing that sampling frequency is 1 kHz, can also be concluded that the number of samples between two points is also the time distance in milliseconds.

Heart rate is also calculated from the time difference between the ECG peaks and is sent to the interface. After the calculation of PTT, a number of 12 values of SPTT and DPTT are collected in an array, averaged and processed to remove possible out of range PTT values.

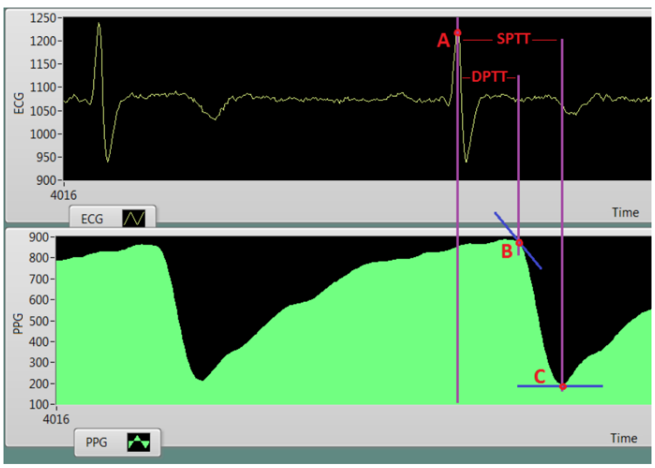


Figure 9. Selected points to calculate PTT values

V. INTERFACE AND COMMUNICATION PROTOCOL

As described in the previous section, in the final version of the wrist-worn sensor system all signal and data processing is to be executed by the microcontroller embedded on the device. The system hardware in use, the H2O device, does not have display included. Tools available were not useful in our case as showing at the same time single values and graphs with different scales is required. A custom Graphical User Interface (GUI) that can communicate with the board and visualize all needed data sent from the board is created.

While plotting, the interface should also be able to show the single values of PTT, heart rate, blood pressure, and before averaging raw PTT values. Another important feature needed was being able to export recorded data in a particular, easy to read format. Exported data are stored for potential data re-processing in the future.

The interface communicates with the board through a serial COM port. There are three plotting windows implemented on the interface, as shown in Figure 10. ECG and PPG are plotted at 250 Hz, ¼ of the frequency of sampling. The frequency of 250Hz is sufficient to plot the complete waveform and it is more efficient for the microcontroller. The window labelled with “PTT” plots last 25 values of systolic PTT (SPTT) and diastolic PTT (DPTT) before being averaged. On the right side of the PTT window, are printed the last 12 values of PTT, which are used to calculate the mode and derive the single value of PTT. Single values of PTT are printed in two textboxes, labelled “DPTT” and “SPTT”.

The second functionality of the interface is the ability to personalise the datasets and export them to excel for further analysis and validation if necessary and to enable the development of the algorithms to establish the parameters affecting BP. Associated with the button labelled “Excel” is a textbox to write the name of the subject. After clicking with a mouse on it, a new window shows a questionnaire for the subject to capture all relevant data such as gender, age, arm span, dietary factors and physical activities. All the data can be exported into a custom excel format for analysis. In the excel file, will also be exported by default all the data from the plotting windows.

This interface was developed to enable the testing and evaluation of our BP monitoring system, but it can also be useful to be used in other applications requiring visualisation of serially transmitted sensor data. It is designed in such a way that if you follow the simple communication protocol, it will plot data from applications that require a similar functionality.

In order to be able to use this interface for other purposes, the communication protocol is described briefly. The general idea of the protocol was to be generic, so it would be easily expanded if required. To do this, for every plot or textboxes values, a specific character is assigned. Data is sent in one single array with all the characters preceding the values. This means that if one needs to plot graphs and print single values simultaneously, using the same COM port, then this interface can be used.

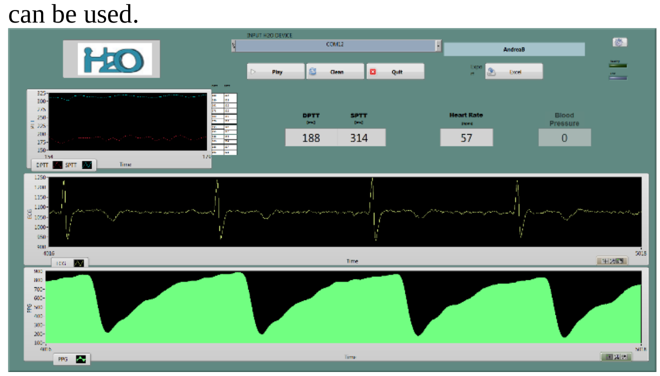


Figure 10. A view of the main window of the interface

Figure 11 illustrates the protocol and the creation of the array that would be accepted by the interface. The order of values per window is not important as the parser in the interface processes separately every string between two commas.

Generic format:

ECG1,	ECG2,	ECGn,	PPG1,	PPG2	PPGn,	PTT1,

…	PTT2	PTTn,	DPTT,	SPTT,	HR,	BP,

Example: 4Hz

E25,	E27,	E20,	E31,	P74,	P86,	P55,	P69,

…	T340,	T344,	T341,	T335,	D212,	S341,	H65,

Figure 11. Generic format and an example of the communication protocol

The string parser starts by reading the array. Once it recognizes the START character, it will look for the value, which should be immediately after the character, and a comma is the end of one value. The recognized value is printed on the respective plot. If the character corresponds to one of the textboxes, the value is shown in a textbox. In case of plots, you expect to have more than one value. In our case, 1000 values every 4 seconds are sent and plotted. The reasons for this size of windows as explained in Section IV are: not delaying the overall execution time in case the algorithm fails to calculate PTT in one window, and generally a 4 second window is sufficient to locate three heart beats. Then, array members are parsed in sequence, printed, and shifted to the left.

If this interface needs to be reused for other purposes, all one has to do is to send a sequence of the values with the corresponding values for each window. An example is given in Figure 11. In this example, 4 samples per second are sent, therefore plotted. The last 3 values in the array, are textbox values, and they are shown in textboxes until the new value arrives.

VI. CLINICAL TRIALS PROTOCOL

Studies mentioned in the Section II show clearly that PTT and BP are related, but different factors, also discussed in Section II, with time are causing changes in this relationship [21]. In order to explore better those factors, the influence they have in BB-PTT relation, and potentially including those findings in BP estimation algorithm, a clinical trial protocol is designed and measurements are carried-on with volunteers. In this section are shown details from the protocol and processed results after measurements.

Before starting with measurements, based on literature, a number of different factors that may affect BP-PTT relation are listed. Based on that, the study is divided into the interventional part and non-interventional one. An interventional study is a type of clinical study in which participants are assigned to groups that receive one or more intervention/treatment so that researchers can evaluate the effects of the interventions on biomedical or health-related outcomes. The assignments are determined by the study's protocol. Participants may receive diagnostic, therapeutic, or other types of interventions. In this study, is has been started with the non-interventional part, which can be completed faster and can provide feedback for the interventional part, if it is needed.

Clinical trials protocol is designed in collaboration with cardiologists, and this research study has received Ethics approval from the Clinical Research Ethics Committee (CREC) of the Cork Teaching Hospitals. The study is also covered by the insurance policies of the University College Cork. The interventional part should go through the same procedure, once it is considered necessary to be carried-on. The process of clinical trials and how data will be applied is described in a flowchart in Figure 12.

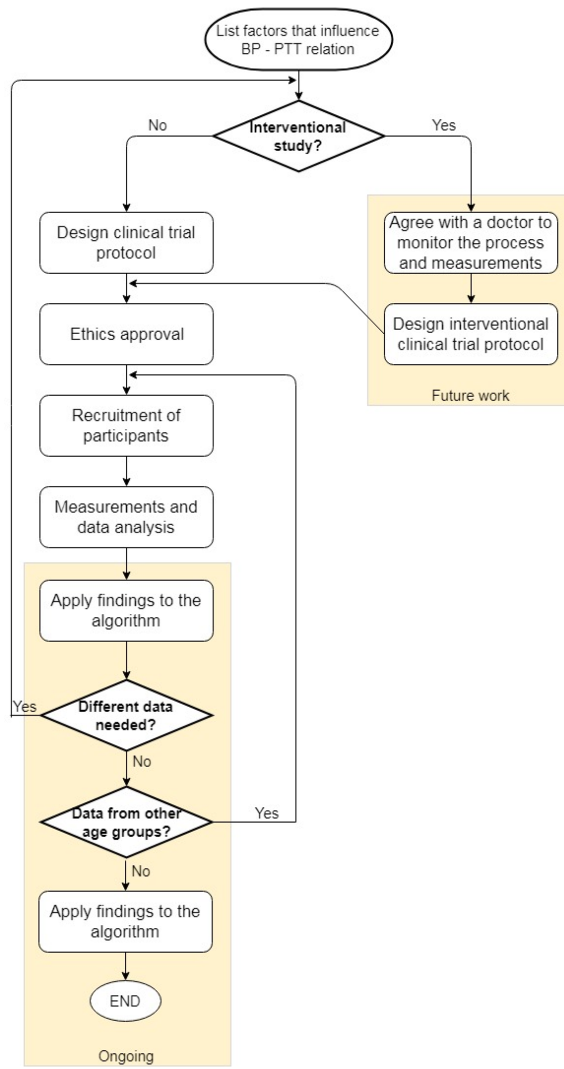


Figure 12. The flowchart of the process of data collection and data usage

During the first phase, 33 participants were recruited. Participants are all between 20 and 35 years old. They were suggested to log all their activities and food they intake from 6 pm of the day prior to the day of measurements and until the last measurement. There are five measurements during one day, and they are organized as follows:

1. In the morning, within 2 hours after waking up
2. Within 30 minutes before lunch
3. Within 1 hour after lunch
4. After 4 pm, before coffee (if participants use coffee)
5. 30 minutes after coffee

The protocol is designed in such a way, that at the conclusion of the experiment we can have datasets which allow us to explore the relationship between BP-PTT as measured in the morning, compared to an evening measurement. In addition we can explore the relationship between these parameters before and after drinking coffee, as well as before and after eating lunch. In this way we can, or see any potential influence of a particular food or different activities on the BP-PTT relationship. After designing the blood pressure estimation algorithm and evaluating the accuracy and results, measurements with other age groups may be useful to undertake if the data from the first trials are useful for the estimation algorithm.

Each measurement took approximately 10 minutes to carry out. The measurements were carried out while participants were in a seated position. The standardised position of the patient during the experimental measurement session is shown in Figure 13.

Figure 13. Set-up of the measurements during clinical trials

As a gold standard, a medically approved BP monitor, BP-801 from Withings, is used to measure blood pressure at the same time as the H2O system developed by the team. This product has received clearance from the Food and Drug Administration (FDA) in the USA and is compliant with European medical device regulations [23]. During the experiment, PPG is recorded continuously using the H2O device. In order to see and record ECG, the user needs to place the index finger of the opposite arm on the electrode to read the differential signal. A closer view of how the watch is worn and the position of the finger is given in Figure 14.

Figure 14. Board with enclosure and worn on the wrist

The procedure is as follows:

1. Sit down, with feet flat on the ground and the back supported
2. Wear the gold standard commercially available BP monitor and take a BP measurement (left arm)
3. Wear the H2O device on the opposite arm with where commercial BP monitor is worn (wrist of right arm)
4. 3-5 minutes after the first BP reading, take the second BP reading using the BP monitor
5. Touch the ECG electrode of the H2O device with the index finger of the opposite arm to close the electrical circuit as described above.
6. Record the ECG signal and calculate PTT for no more than 2 minutes

Before taking the measurement in the morning, for each patient, we also measured arm span, height, the weight of the participant, as well as the room temperature at the time of measurement. To ensure the most accurate results, and to make the participants feel more comfortable and relaxed with the air pressure cuff, two measurements are carried-on within one session.

In general it was observed that the blood pressure dropped in the second measurement, after sitting still for 3-5 minutes. As a result, based on our measurements and suggestions from doctors, the second measurement is the one chosen to be the more accurate result, and was registered as a BP value for one session. In some cases, first and second measurement were significantly different from each other, and a third measurement need to be taken and recorded.

Processed results of measurements are presented in Table I. Three records of diastolic BP-PTT relation are discarded because the algorithm was unable to determine the value of DPTT as the PPG signal was not clear. First column of Table I shows four categories of morning and evening differences of BP and PTT. The second and the last column show the number of subjects in each category for systolic BP and diastolic BP respectively.

Table I. Morning and Evening relation of BP-PTT

Differences between morning and evening in BP and PTT	Number of subjects [Systolic]	Number of subjects [Diastolic]
Increased BP – Increased PTT	2	6
Increased BP – Decreased PTT	14	3
Decreased BP – Increased PTT	6	10
Decreased BP – Decreased PTT	11	11
Total	33	30

VII. PHASE 2 DESIGN STRATEGY OF THE DEVICE

Currently, a second generation system is under development incorporating all the necessary system functionalities to enable physiological monitoring. Designed around a "Flex/Rigid" system design methodology so as to enable a highly miniaturised robust prototype, the system functionalities include sensors for ECG (Electrocardiography), SpO2 (Oxygen Saturation) levels, Heart rate, Blood pressure and Activity/Motion tracking. Communications modalities include Ultra Wide Band (UWB) for both communications and location tracking/ranging as well as Bluetooth Low Energy (BTLE) to facilitate communications between the smartwatch and any mobile device belonging to the user for additional data analytics and visualisation.

Figure 15. Details of the Phase 2 design of the device

In Figure 15 is shown how the physiological sensors are placed in the new device design. There is a difference in the design of the PPG sensor in this version compared to the PCB used in the first phase of experiments described above. Two photodiodes are used instead of only one as in the previous version and also the SpO2 sensor is designed separately in which a dedicated circuit, photodiode, red LED and infrared LED are used. The complete device design is presented in Figure 16.

Figure 16. Design of the final device (future work)

VIII. CONCLUSIONS AND FUTURE WORK

There is no doubt that a BP monitor that would be accurate, reliable, cuffless, and comfortable with easy to carry out frequent or even continuous measurements would be priceless for clinical diagnosis of cardiovascular illness diagnosis. For decades, research has been carried out to achieve this goal. Pulse transit time (PTT) seems to be the most promising method to achieve it, based on literature. But until now, one of the challenges to be addressed is in the development of an appropriate data acquisition system to provide the necessary data sets for such a system. The main challenges are: vessel's properties changing from person to person, vessel's properties can be changed by factors within body or ambient conditions and clear data acquisition from comfortable wearable sensors.

The WSN group at Tyndall is currently developing such a system. In this paper, we are showing practical results of the first phase of the work in progress. The main focus of this stage is to deliver a complete system that allows proceeding with a deeper evaluation of the BP-PTT relation. The first part of the study is focused on the evaluation of the computing hardware, sensors, custom board design and data acquisition. Initial results of evaluating the integrated ECG sensor and PPG sensor are shown during the first three sections, including the new PPG sensor design for the wrist.

In Sections IV and V are presented results on developing the required algorithms and software to calculate PTT from biosignals and data visualisation.

Finally, we have used the new PPG sensor designed, the Tyndall custom PCB board and developed algorithms to run measurements with participants in clinical validation trials with clinical partners and continue the development of the required processing and algorithms to provide BP measurements from PTT measurements. A list of factors that affects blood pressure significantly has been created, and measurements are carried out to try to quantify the influence they have. These measurements could enable the development of the required algorithms that would relate BP and PTT, which is the main part of the complete study to be reported in subsequent publications.

ACKNOWLEDGMENT

These research is a part of the H2O (Human to Objects) project funded by the European Union under the CATRENE program. This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund under Grant Number 13/RC/2077.

REFERENCES

[1] A. Hajzeraj, M. Belcastro, D. Alfieri, and B. O'Flynn, "Evaluation Phase of a Novel Blood Pressure Monitor Device", in Global Health 2017, IARIA, Editor. 2017: Barcelona, Spain. pp. 14-19.

[2] (EHN), E.H.N., "European Cardiovascular Disease Statistics 2017".

[3] J.-M.R. Dilpreet Buxi, Mehmet Rasit Yuce, "A Survey on Signals and Systems in Ambulatory Blood Pressure Monitoring Using Pulse Transit Time", 2015.

[4. R. Mukkamala, J.O. Hahn, O.T. Inan, L.K. Mestha, C.S. Kim, H. Töreyin, and S. Kyal, "Toward Ubiquitous Blood Pressure Monitoring Via Pulse Transit Time: Theory and Practice", IEEE Transactions on Biomedical Engineering, 2015. 62(8): pp. 1879-1901.

[5] L. Peter, N. Noury, and M. Cerny, "A Review of Methods for Non-Invasive and Continuous Blood Pressure Monitoring: Pulse Transit Time Method Is Promising?". IRBM, 2014. 35(5): pp. 271-282.

[6] B.M. McCarthy, B.O. Flynn, and A. Mathewson, "An Investigation of Pulse Transit Time as a Non-Invasive Blood Pressure Measurement Method", Journal of Physics: Conference Series, 2011. 307(1): pp. 012060.

[7] R.D. Brook, A.B. Weder, and S. Rajagopalan, ""Environmental Hypertensionology" the Effects of Environmental Factors on Blood Pressure in Clinical Practice and Research", J Clin Hypertens (Greenwich), 2011. 13(11): pp. 836-842.

[8] M.S. Ganio, R.M. Brothers, S. Shibata, J.L. Hastings, and C.G. Crandall, "Effect of Passive Heat Stress on Arterial Stiffness", Experimental physiology, 2011. 96(9): pp. 919-926.

[9] T. Frawleyand T.B. Bunton, "Effect of Temperature on Pulse Wave Velocity and Arterial Compliance", Journal of Undergraduate Research in Physics, 2012.

[10] M. Jehn, L.J. Appel, F.M. Sacks, and E.R. Miller, 3rd, "The Effect of Ambient Temperature and Barometric Pressure on Ambulatory Blood Pressure Variability", Am J Hypertens, 2002. 15(11): pp. 941-945.

[11] S. Lluberas, D. Bia, Y. Zócalo, M. Zabalza, C. Etchart, and R. Armentano, "Sleep-Wakefulness Variations in Arterial Stiffness: Assessment Using Ambulatory Recording of Arterial Pulse Transit Time", Revista Española de Cardiología (English Edition), 2008. 61(01): pp. 41-48.

[12] D. Bia, S. Lluberas, Y. Zócalo, C. Etchart, M. Zabalza, and R.L. Armentano, "Circadian Pattern and Night-Day Variations in Human Arterial Stiffness: Assessment Using Ambulatory Recording of Arterial Pressure and Pulse Transit Time", in Iv Latin American Congress on Biomedical Engineering 2007, Bioengineering Solutions for Latin America Health: September 24th–28th, 2007 Margarita Island, Venezuela, C. Müller-Karger, S. Wong, and A. La Cruz, Editors. 2008, Springer Berlin Heidelberg: Berlin, Heidelberg. pp. 82-86.

[13] S.-H. Liu, D.-C. Cheng, and C.-H. Su, "A Cuffless Blood Pressure Measurement Based on the Impedance Plethysmography Technique", Sensors, 2017. 17(5): pp. 1176.

[14] C. Byeong Cheol, L. Hee Jeong, Y. Soo Young, J. Dong Keun, K. Gi Ryon, K. Kwang Nyon, and J. Gye Rock. "Evaluation of Arterial Compliance on Pulse Transit Time Using Photoplethysmography", in 30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON 2004. 2004.

[15] "Stm32 Ultra Low Power Mcus.", [Retrieved 26/02/2018]; Available from: http://www.st.com/en/microcontrollers/stm32-ultra-low-power-mcus.html?querycriteria=productId=SC2157.

[16] Ltd, P.S., "Epic Sensor Applications Guidebook", pp. 17-22.

[17] "Afe4490 Integrated Analog Front-End for Pulse Oximeters.", [Retrieved 26/02/2018]; Available from: http://www.ti.com/product/AFE4490/datasheet.

[18] T. Seeberg, J. Orr, H. Austad, M. Roed, S. Dalgard, D. Houghton, D. Jones, and F. Strisland, "A Novel Method for Continuous, Non-Invasive, Cuff-Less Measurement of Blood Pressure: Evaluation in Patients with Non-Alcoholic Fatty Liver Disease", IEEE Transactions on Biomedical Engineering, 2016. PP(99): pp. 1-1.

[19] "Ad8232 Single-Lead, Heart Rate Monitor Front End.", [Retrieved 26/02/2018]; Available from: http://www.analog.com/media/en/technical-documentation/data-sheets/AD8232.pdf.

[20] S. Hey, A. Gharbi, B.v. Haaren, K. Walter, N. König, and S. Löffler. "Continuous Noninvasive Pulse Transit Time Measurement for Psycho-Physiological Stress Monitoring", in 2009 International Conference on eHealth, Telemedicine, and Social Medicine. 2009.

[21. M. Sharma, K. Barbosa, V. Ho, D. Griggs, T. Ghirmai, S. Krishnan, T. Hsiai, J.-C. Chiao, and H. Cao, "Cuff-Less and Continuous Blood Pressure Monitoring: A Methodological Review", Technologies, 2017. 5(2): pp. 21.

[22] A. Esmaili, M. Kachuee, and M. Shabany, "Nonlinear Cuffless Blood Pressure Estimation of Healthy Subjects Using Pulse Transit Time and Arrival Time", IEEE Transactions on Instrumentation and Measurement, 2017. 66(12): pp. 3299-3308.

[23] "Withings Wireless Blood Pressure Monitor", [Retrieved 26/02/2018]; Available from: https://health.nokia.com/ie/en/blood-pressure-monitor.

A CTS2 Based Terminology Service for Managing Semantic Interoperability in the Italian Federated Electronic Health Record

Elena Cardillo and Maria Teresa Chiaravalloti

Institute of Informatics and Telematics
National Research Council, IIT - CNR
Rende, Italy
Email: elena.cardillo@iit.cnr.it, maria.chiaravalloti@iit.cnr.it

***Abstract* - Semantic interoperability is crucial for data and document exchanges between and among Healthcare Information Systems, and to assure efficient communication among the different domain experts involved in the patient process of care. In this context, integrated terminology services offer the chance to manage clinical code systems, both standard and local, as well as value sets, through a series of functionalities such as searching, querying, import/export, cross-mapping, etc. This paper describes the approach used for the implementation of an integrated terminology service, namely *Servizio Terminologico Integrato*, in the Italian Federated Electronic Health Record setting. The service is based on Common Terminology Service Release 2 by Health Level 7, which is the main referenced standard in the domain and has the aim to support domain experts and healthcare organizations in managing medical terminologies for ensuring semantic interoperability. Preliminary tests show promising results and the usefulness of such a service, above all for the retrieval of coding systems mapping and updates. The main output of this paper is the implementation in the Italian healthcare setting of an open source, standard-based and bilingual integrated terminology service to support professionals and stakeholders in sharing, querying and maintaining official and up-to-date terminological artifacts.**

Keywords* - *Code Systems; Semantic Interoperability; Terminology Services; Common Terminology Service Release 2 (CTS2); Healthcare.

I. Introduction

Nowadays interoperability is the word most commonly associated to e-health services. It essentially means that different systems are able to communicate among each other, exchange data, and, above all, reuse them. The general aim is ensuring a worldwide availability of information at the right time and place, in order to deliver better clinical services and improve healthcare. Interoperability is a required function for the proper use of Electronic Health Record (EHR) systems, which remain simple data containers if they do not have the chance to dialogue using the same language. Standardized coding systems are the lingua franca of medical data, as they allow to uniquely identify the same concept regardless of the language, synonyms or local names that could be used to refer them. The advantages of standards are commonly recognized and their usefulness increases over time as they are employed in numerous health-related Information Technology applications. Nonetheless, their use is not always as easy as it may appear and health professionals often complain about the lack of adequate support systems.

Managing clinical terminologies is not only a matter of making them available to the users, but their management needs to include further functions to offer a complete plethora of services allowing a meaningful use of standards. To pursue this aim, there is the need for a standard protocol to manage terminology standards in the same way across multiple healthcare facilities. This role is covered by integrated terminology services, which offer the possibility to interact with terminologies according to a series of standardized functionalities, such as research, hierarchical tree navigation, structured query, cross mapping.

The objective of this paper, which is an invited extension of [1], is to describe the approach used for designing and developing an integrated terminology service, called STI (acronym of the Italian *Servizio Terminologico Integrato*), to provide support in the use of the clinical coding systems in the Italian Healthcare setting, in particular to ease the main functionalities that users, including both Regions and health facilities, and domain experts, could need in managing terminologies and codes in the documents required in the national federate EHR infrastructure (corresponding to the Italian acronym FSE, meaning "Fascicolo Sanitario Elettronico"), realized by the Italian National Research Council (CNR) in accordance with the Digital Italy Agency (AgID). This infrastructure allows the exchange of clinical documents among the regional EHR systems [2].

In this framework, semantic interoperability is a non-trivial issue, especially because, over time, regional and local coding systems and habits have proliferated. In [3] authors provide a detailed overview of the main issues for the semantic interoperability implementation in Italy. To provide general guidelines on the use of the FSE, the Prime Minister Decree No. 178/2015 [4] was issued, also making some medical terminologies mandatory, and detailing their use in the two kinds of documents (Patient Summary and Laboratory Report) included in the minimum unit expected in the FSE first implementation phase.

To develop the STI terminology service, the standard protocol Common Terminology Services Rel. 2 (CTS2) [5] by Health Level 7 (HL7) was tested.

The paper is organized as follows. Section II gives an overview of the state of the art on semantic interoperability and the main features of CTS2. Section III describes the material used within the Italian implementation of the STI system and Section IV addresses the content approach. Section V shows some preliminary results. Discussion and conclusions in Section VI close the paper.

II. Background

A. Semantic Interoperability: projects and initiatives

The adjective *semantic* conveys the deep meaning of interoperability as it overcomes lexical and syntactical issues to deal with the meaning of the exchanged information. The best EHR system would be useless without semantic interoperability, as it could not unambiguously interpret data received from other systems. In fact, Semantic Interoperability in healthcare is defined as "the ability of a healthcare system to share information and have that information properly interpreted by the receiving system in the same sense as intended by the transmitting system" [6].

Projects and initiatives address the semantic interoperability issue trying to propose effective solutions to solve it. Regarding European Community (EU) initiatives and projects, it is worth mentioning the FP7 project *Semantic Interoperability for Health Network*, whose main aim was the implementation of the necessary infrastructure and governance to allow sustainable semantic interoperability of clinical and biomedical knowledge at the European level [7]. Furthermore, the EHR4CR project [8] dealt with the development of a semantic interoperability service platform, which includes a mediation model for multiple standards integration and harmonization. It was tested in 11 EHR systems of 5 EU Countries. Finally, the Trillium Bridge and Trillium Bridge II projects involve EU Countries and US for the creation of a shared model of an International Patient Summary (IPS), to improve semantic interoperability of e-health systems beyond EU borders [9].

Also, international standards organizations proposed protocols for semantic interoperability. The main one is the CTS2 standard proposed in the Healthcare Service Specification Program (HSSP), a joint HL7 and Object Management Group (OMG) initiative [10], where HL7 provided the requirements as a Service Functional Model (SFM), and OMG developed the formal specification of the standard. CTS2 is a cohesive model and specification for representing, accessing, querying, exchanging and updating terminological resources (e.g., Code Systems, Value Sets, Mappings), built on the RESTful (Representational state transfer) Architectural Style. More recently, HL7 proposed the Fast Healthcare Interoperability Resources (FHIR) Specification [11], another standard for exchanging healthcare information electronically, which, compared to the previous HL7 standards, is more consistent and easy to implement, thanks to its built-in extension mechanism to cover the needed content. In fact, specific use cases can be implemented by combining resources together through the use of resource references.

In the literature, different initiatives aimed at developing terminology integration platforms or services were launched. Initial studies and applications focused, for example, on the use of the Unified Medical Language System (UMLS) Metathesaurus, developed by the US National Library of Medicine [12], which includes more than 100 biomedical vocabularies integrated on the basis of a common Semantic Network and mapped among them. Researchers use UMLS to create knowledge-based representation for controlled terminologies of clinical information and to extract and validate semantic relationships. Particularly relevant are also the HETOP terminology service [13] and the LexGrid initiative [14]. The former includes cross lingual multi-terminological mappings on a semantic basis, while the latter promotes the use of common terminology models to accommodate multiple vocabulary and ontology distribution formats, as well as the support of multiple data storing for federated vocabulary distribution. On the contrary, PyMedTermino [15] implements a generic model and has a separate database for each terminology to be included into the service. This allows for better preservation of the original architecture of each terminology and the easy integration of post coordinated terminologies. PyMedTermino supports advanced operations, such as derived queries, on the concepts set, which are not usually feasible into other terminology services. Nonetheless, its limits are: exclusive Python implementation, no OWL ontologies link, low quality of the UMLS mapping. Luna et al. [16] describe the implementation of an interinstitutional and transnational Remote Terminology Service, which uses local interface terminologies (in the form of a thesaurus) mapped to reference terminologies (such as the Systematized Nomenclature of Medicine - Clinical Terms, namely SNOMED-CT) to allow users to enter free text diagnosis and procedures and get back coded content. It is also possible to select preferred reference terminology among the standardized clinical classification systems, as the service uses the SNOMED-CT cross matchings. The service reveals its utility in polysemy, synonyms and acronyms management, but it is only available for Spanish.

In the last few years, much effort has been spent on the application of the abovementioned CTS2 standard to develop terminology services, such as that realized by the Mayo Clinic Informatics, which is the most internationally relevant [10]. D2Refine Workbench platform, for example, aimed at standardizing and harmonizing clinical study data dictionaries [17]. Focused on laboratory catalogues, the experience of the Partners HealthCare System of Boston, applies the CTS2 *Upper Level Class Model* to represent and harmonize the structure of both local laboratory order dictionaries and reference terminologies [18]. Peterson et al. presented, instead, a design user-centered approach, based on the use of Extraction, Transformation and Loading (ETL) procedures in CTS2-based terminology services [19]. The main advantages of this service are: i) adaptability, ii) interoperability, because of the numerous standard vocabularies included, iii) usability, since it is focused on users' needs. In the wake of these projects, we propose a multi-layer CTS2 implementation that is not only based on ETL procedures, but also allows mapping (and their validation) between local dictionaries and standardized code systems, including semantic enrichment through external ontological references.

Interesting applications of CTS2 can also be found in the European context, where the main implementation is the Standard Terminology Services (STS) provided by the French non-profit development standards and services organization PHAST [20]. Other implementations are the following: the Austrian national patient health record ELGA, where all relevant clinical terminologies are provided through a CTS2-conformant terminology server [21]; and the Terminology Server, realized by the University of Applied Science of Dortmund, which also offers a collaboration environment to develop terminologies in a team [22]. Finally, in Italy, the existing implementations of CTS2-based terminology services are proprietary solutions, i.e., the Distributed Terminology Assets Management system (DITAM), a software designed by Codices-Noemalife that offers standard terminology services through a federate architecture [23]; and the HQuantum (a spin-off of the University of Genova) with the Health Terminology Service (HTS) [24], which is especially focused on the management and integration of local laboratory data through the LOINC standard. These two solutions were evaluated as non-fitting for the purpose of our project because they are subject to license, while the FSE project required an open and reusable solution. Furthermore, they were, at the time of the initial efforts on this activity, only partially developed and tested, and this would have implied a lot of customization effort.

B. *CTS2 HL7 overview*

As stated in the ANSI/HL7 V3 CTS R2-2015 standard [5], "the HL7 Common Terminology Services (HL7 CTS) is an API specification that is intended to describe the basic functionalities that will be needed by HL7 Version 3 software implementations to query and access terminological content. It is specified as an Application Programming Interface (API) rather than a set of data structures to enable a wide variety of terminological content to be integrated within the HL7 Version 3 messaging framework without the need for significant migration or rewrite". The standard, currently, consists of:

- CTS2 Normative Edition v1.0 and the Service Functional Model (SFM), which serve as Functional Requirements Documents, defining the capabilities, responsibilities, inputs, outputs, expected behavior and a set of core functionalities to support the management, maintenance, and interaction with ontologies and medical vocabulary systems.
- CTS2 Technical Specification, which serves as a technical specification document to define the precise API interface specifications for CTS2 implementation compliance in Simple Object Access Protocol (SOAP).

The CTS2 Information Model specifies the structural definition, attributes and associations of Resources common to structured terminologies such as Code Systems, Binding Domains and Value Sets. In particular, in the CTS2 Service Functional Model the following semantic entities are distinguished (see [5] for details):

- *Code Systems* – defined as a collection of codes with associated designations and meanings. To meet the requirements of a code system as defined by HL7, a given code must refer to one and only one meaning within the code system. A code system in CTS2 Functional Model can be of different types:
  - Set of codes – a flat list of code/term pairs;
  - List of terms – a set of words (or tokens), each one associated with an unambiguous definition, if used in a specific context;
  - Thesaurus – a controlled vocabulary, i.e., a (semi-) structured set of words or locutions (structured according to semantic relations) whose aim is to allow and facilitate information retrieval and navigation;
  - Classification system – a set of hierarchically organized categories or classes used to organize and aggregate data in a specific domain of knowledge or for a specific aim;
  - Ontology – a formal representation or a conceptualization of a given domain of reality.
- *Value Sets* – defined as uniquely identifiable sets of valid concept representations (Codes), where any concept representation can be tested to determine whether or not it is a member of the value set. More practically, they are subsets of codes, classes or descriptions of entities selected from one or more code systems and grouped for a specific purpose (e.g., a value set of units of measures, a value set of anatomical parts, etc.).
- *Maps* – a set of formal rules that permits concepts (EntityDescription) transcoding from a starting value set or code system to a second one.
- *Concept Domains* – metasyntactic variables in database models, messages or other representative forms.
- *Entity Description* – the definition of a concept in the context of a particular terminological system.
- *Association* – the semantic statement (subject – predicate – object) that associates entities within the same code systems or in different code systems.

The Computational Model specifies the services descriptions and interfaces needed to access and maintain structured terminologies. The main CTS2 profiles and functionalities are:

a) *Search/Query Profile:* i.e., “read-only” functionalities and availability of the resources to the client. It includes: reading of a resource, a code or a concept; browsing or visualization of the tree of a resource; the download of a resource.
b) *Terminology Administration Profile*: i.e., “write” administrative functionalities. It includes: import of a resource; creation of mappings between the imported resources; the possibility to use updates and notifications.
c) *Terminology Authoring Profile*: i.e., “read-write” functionalities. They are intended for an application used by specialized users (e.g., translators, or domain experts) to create and maintain terminological resources.

More specifically:

- *Read* – is the direct access to the contents of a resource via URI, local identifier or, where applicable, a combination of an abstract resource identifier and version tag (e.g., LOINC/Current version).
- *Query* – is the ability to access, combine and filter lists of resources based on their content and user context.
- *Import and Export* – correspond to the ability to import external content into the service and/or export the contents of the service in different formats.
- *Update* – is the ability to validate load sets of changes into the service that updates its content.
- *History* – is the ability to determine what changes occurred over stated periods of time.
- *Maintenance* – is the ability to create and commit sets of changes.
- *Temporal* – is the ability to query on the state of the service at a given point in the past (or in the future).
- *Specialized* – are service specific functions such as the association reasoning services, the map entry services and the resolved value set services.

The CTS2 Development Framework is a development kit for rapidly creating CTS2 compliant applications. It allows users to create plugins, which may be loaded into the Development Framework to provide REST Web Services that use CTS2 compliant paths and model objects. Since it is plugin based, users are only required to implement the functionality that is exclusive to their environment. Thus, CTS2 Development Framework provides all the infrastructures and utilities to help users create plugins. Given the short time available to develop the service, in this work, we reused the mentioned CTS2 Development Framework toolkit provided by Mayo Clinic Informatics [10] available from Github, which is useful for rapidly creating CTS2 compliant applications, and, at the moment, it is recognized as the most complete and documented. Furthermore, the community that uses the Development Framework is wide and quite reactive.

III. MATERIAL

The terminology service was designed taking into account both the CTS2 main functionalities requirements and the structures of the medical coding systems required by the law. More specifically, the standards to be used are:

- ICD-9-CM (*International Classification of Diseases – 9th revision – Clinical Modifications*) [25], developed by the World Health Organization (WHO): its use is required in the Patient Summary for coding relevant and chronic diseases, in Prescriptions and in Discharge Letters for coding diagnoses;
- ATC (*Anatomical Therapeutic Chemical Classification*) [26], developed and maintained by the WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health: its use is required in the Patient Summary for coding adverse reactions to food and medication, the medication plan, and vaccinations;
- AIC (*Autorizzazione all’Immissione in Commercio*) [27], developed and maintained by the Italian Medicines Agency (AIFA): as for ATC, its use is required in the Patient Summary for coding adverse reactions to medication, the medication plan, and vaccinations;
- LOINC (*Logical Observation Identifiers Names and Codes*) [28], developed by the Regenstrief Institute Inc. in Indianapolis: its use is required in the Laboratory Report for coding performed tests and their specialty or class. LOINC codes are also used for identifying the clinical document type, when it is structured according to the HL7 Clinical Document Architecture rel.2 (CDA2).

As each resource has a different structure, the most suitable solution was integrating them into the STI Knowledge Base allowing the correct visualization and searching into each of them. ICD-9-CM, for example, is a classification, which has a hierarchical tree structure so in the visualization it needed the use of indentations and expandable/collapsible branches for navigating the tree; LOINC, instead, is more like a nomenclature, without any hierarchical structure (codes are progressive and not informative). Furthermore, each LOINC code has associated numerous information to be visualized, which are discriminative in choosing a code rather than another, so it was necessary to realize a personalized form to access LOINC code details (e.g., System, Scale, Method, etc.).

As further explained in Section IV.B, a great deal of effort was spent on the collection of the different versions of each standard and on the re-structuring of the available files according to the CTS2 concept model, to integrate them in the STI Knowledge Base. In addition to the standard terminologies required by the law, some other resources, such as value sets and local files mapped to the code systems, were integrated into STI. The first type includes files of synonyms of LOINC and ICD-9-CM terms, which could be used as further research items to find a specific code. For the second type, local files mapped to the code systems were included in STI because they are useful as both basis of comparison for who is working on the same type of mapping and collector of local synonyms of the official terms used in the standard clinical terminologies. At the moment, the service includes:

- 106 local tests from 6 different laboratories in the Umbria Region mapped to LOINC 2.54;
- 154 local tests from a laboratory in the Campania Region mapped to LOINC 2.34;
- 1071 local tests from 6 different laboratories in the Calabria region mapped to LOINC 2.54.

IV. APPROACH

The proposed solution consists of a standard-based and web-based distributed software infrastructure, which was required to be designed open and extendable. It aims to support the production, integration, maintenance, and use of the terminological resources according to the CTS2 protocol. To design and develop the terminology service, an *Agile* methodology was applied. This led to an iterative development of the system functionalities, starting from the core ones and continuing with further iterations in the process of analysis and development. Each iteration and progress in the design and development of the functionalities were submitted to tests by terminology and domain experts.

A. STI Architecture

The STI Architecture (Figure 1) was designed and realized by using Full Open Source integrated components:

- Liferay 6.2. CE [29] as environment to create the Web Application. It manages simultaneous user accesses, content versioning and classification. The platform functionalities were realized through the development of appropriate portlet allowing the management of: i) search and visualization of code systems; ii) administration management of import and elimination of code systems.
- Kettle (Pentaho Data Integration) [30], used to realize ETL procedures for data integration during data migration from different Database Management Systems. ETL procedures include: i) heterogeneous data aggregation; ii) data transport and transformation, by performing data cleaning operations, or scheduled-based data storing, on the destination database. ETL procedures are mostly used in the construction and population of the Knowledge Base.
- Virtuoso Open Source Edition [31], developed by OpenLink, used for the management of ontologies and data in RDF. RDF data can be queried through SPARQL endpoints, to facilitate the connection with structured dataset derived from other sources.
- CKAN [32], for the management and publication of Open Data. This open source software allows for cataloging datasets and describes them across a range of metadata that, on the one hand, help users to navigate through information, and on the other hand, facilitate indexing of the same datasets on search engines. In the present work CKAN is useful to export data (i.e., resources in STI) in the Open Data format and to publish them on open data platforms.

Strengths of this architecture and implementation are: i) all the components are open source; ii) it is scalable, modular and easy to maintain; iii) it is installable on open environment without the need for a license.

B. The STI Knowledge Base

The implementation of the STI Knowledge Base started with the integration of the basic elements, represented by the code systems. They were processed by ETL procedures, in order to enrich them with knowledge derived from external services. The modeling of the basic entities contained in these code systems (i.e., medical concepts) was made through Porting on the Database.

The definition of the STI Knowledge Base was based on four application layers: 1) the Data layer; 2) the Integration layer; 3) the Semantic layer (or Interoperability layer); and 4) the Presentation layer.

1) The Data layer: it is represented by the CTS2 Conceptual Model, for the representation of the different types of resources and semantic relationships, and by a relational database, containing the useful facilities to integrate the resources, in particular code systems, in compliance with CTS2. Each concept of the code systems represents the basic entity that composes the atomic information of the conceptual model and is classified according to the structure defined in the HL7 standard, in XML format. In the STI Knowledge Base, the main code systems and terminology standards, described in Section III, are included in different versions after a readaptation of their structure to the CTS2 model, but at the same time, maintaining their specifications. In particular, versions of the code systems included in the STI Knowledge Base are:

- ICD-9-CM Italian 2007 version, counting more than 16,000 codes. Since the official CSV file distributed by the Ministry of Health is incomplete (it includes only the hierarchical structure, thus the codes and the label of the diagnoses), it was necessary to integrate it. To this aim, we reused an ontological version of the system that was built for another project, where each ICD-9-CM code has multiple associated information: i) official description, ii) alternative descriptions, i.e., synonyms, iii) inclusions criteria, iv) exclusions criteria, v) information about the coding of a primary diagnosis vi) information about the coding of additional diagnoses, vii) further notes. In order to provide access to this additional information, a customized form for ICD-9-CM code details was built.

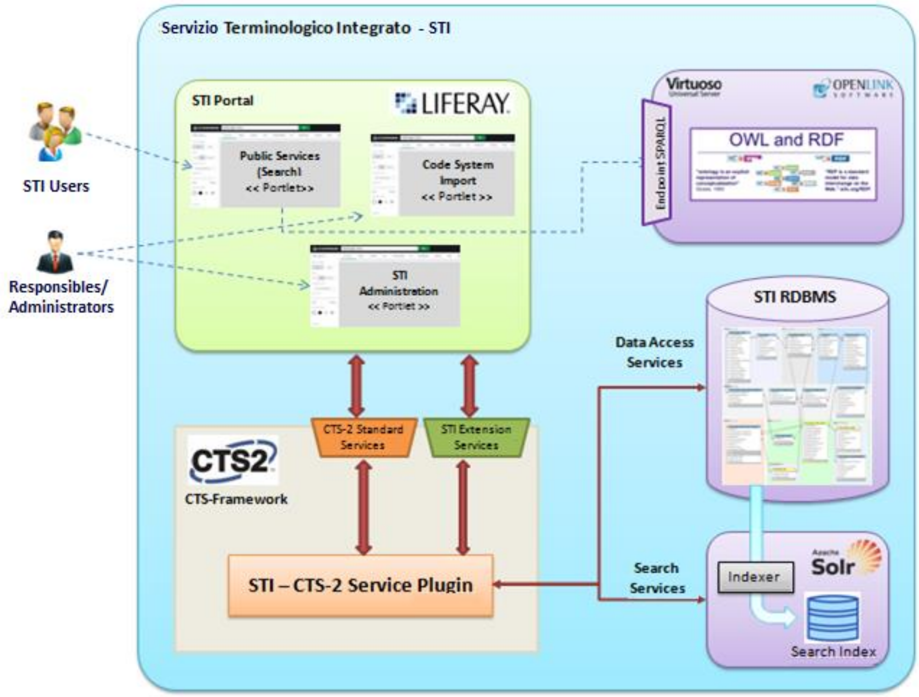


Figure 1. STI Architecture – Deployment Diagram

- LOINC Italian and English versions 2.34 (required by the cited Decree), 2.52, 2.54, 2.56, 2.58, 2.61 and 2.63 (it includes the latest updates of the system, released in December 2017). See Table I for details about LOINC terms. In order to correctly integrate LOINC in the STI Knowledge Base, we needed to upload, for each version, three CSV files: i) the Italian DB, LOINC_IT (it has limited number of fields associated to each code, with respect to the English version); ii) the English database, LOINC_DB (whose structure and fields changed several times over successive versions), and iii) the file including the changes of the mapping codes from one version to another, named *Map_to*. Regarding the LOINC_DB, it was necessary to make all the versions compliant with the structure of the last updates (v. 2.63) and to align the CSV of the Italian version to the same structure.
- AIC January 2017 version (the latest available updates on the AIFA website [27] at this time, including more than 18,000 medicines codes). More specifically, AIC related files are published on the AIFA website as separate files according to the type of drugs. In particular, there are four different CSV files: i) *equivalent_medicines* file, which includes for each AIC code the mapping to the active ingredient and thus the corresponding ATC code; ii) *Class_A_medicines*, including essential medicines and those for chronic illnesses (some of them can also appear in the equivalent medicines file); iii) *Class_H_medicines*, including medicines used only in hospital facilities, which can therefore not be sold in public pharmacies (some of them can also appear in the equivalent medicines file); and iv) *Class_C_medicines*, including drugs that are not licensed by the Italian National Health Service and are therefore to be paid by patients (all AIC codes in this file are mapped to the corresponding ATC code). These four files were separately integrated into the STI Knowledge Base and ETL processes were used to clean and normalize data, in order to avoid concepts overlap.
- ATC Italian 2014 version (the latest one freely available at this time), which counts about 5,000 codes. As for ICD-9-CM, access and navigation of the ATC classification tree was provided.

As AIC and ATC cover the same semantic area, a cross mapping file is constantly updated and available, but not in the form of a unique CSV file with 1-1 mappings. As mentioned above, in fact, only two of the four collected files contain mappings to ATC. In order to create a complete 1-1 mappings file, ETL processes were trained to perform mapping extraction processes from the data into the Knowledge Base. Figure 2 shows the ETL process related to the import of AIC and ATC and the extraction of their mappings.

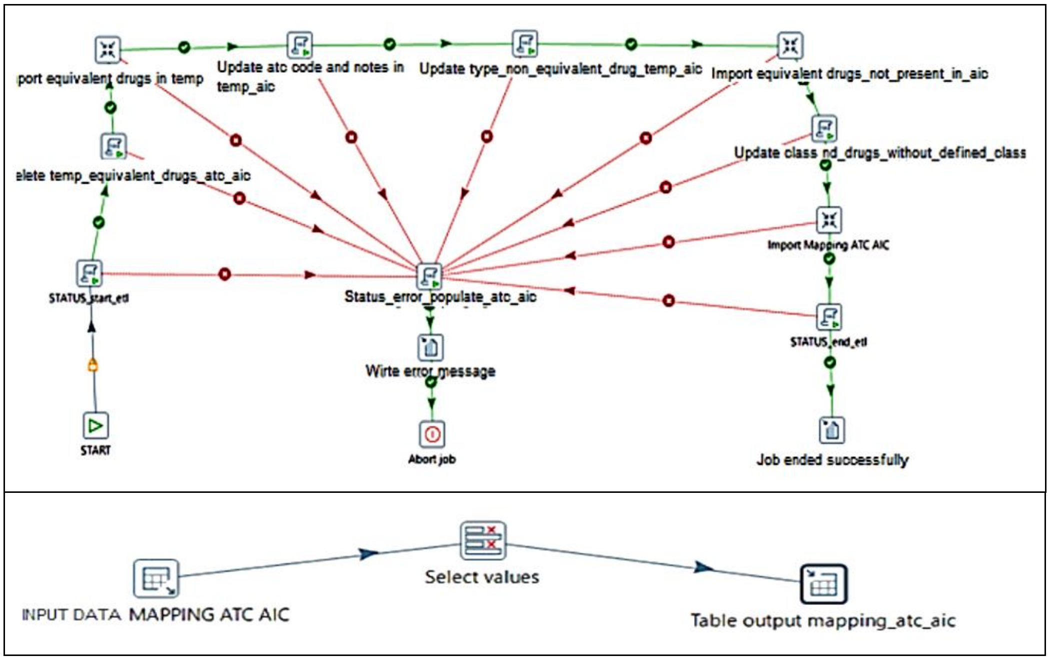


Figure 2. Mappings extraction via ETL: AIC – ATC example

As can be observed, the main job is MAPPING ATC-AIC. From the previous ETL job, expanding the transformation "Import Mapping ATC AIC" we obtain the transformation to import mapping data extracted from the Database (4 files of AIC and ATC). In particular, where an explicit mapping in the different CSV files was not available, it was created by matching the active ingredient of the medicine and by querying external resources (i.e., the AIFA drugs database) in case of ambiguities (i.e., multiple AIC codes associated to one active ingredient), the status changes in AIC, and ATC code updates (considered that we did not use the latest version of the system in our Knowledge Base). The final mapping file establishes mappings from one ATC code to multiple AIC codes, in fact AIC uniquely identifies branded medications while ATC encodes the medication active ingredient. This file was stored as a Mapping resource into the STI in order to give the chance to have a cross reference between the two code systems. In the Semantic layer, the basic information included in the code systems is enriched by semantic content and correlations derived from the Integration layer. Table I gives some statistics about the content integrated in STI Knowledge Base. As observable, not all the code systems are mapped, except for the AIC – ATC, and some local catalogues mapped to LOINC. Since we did not find official and available mappings between the other resources, we provided, as explained in Section V, a functionality to edit mappings directly on the STI platform, under specific permissions and subjected to validation by administration users.

TABLE I. STI KNOWLEDGE BASE STATISTICS

Resources	***Version***	***N. of Concepts (En)***	***N. of Concepts (It)***
LOINC	2.34	+ 60,000	43,152
LOINC	2.52	+ 72,000	58,045
LOINC	2.54	+ 73,000	61,419
LOINC	2.56	~ 80,000	60,837
LOINC	2.58	+ 80,000	63,367
LOINC	2.61	~ 85,000	63,367
LOINC	2.63	+ 86,500	66,204
ICD-9-CM	2007	16,100	16,100
ATC	2014	-	5,530
AIC	January 2017	-	18,309
Umbria_laboratories tests Catalogues - LOINC 2.54 Mapping	2016	-	106
Campania_laboratory _tests Catalogue - LOINC 2.34 Mapping	2012	-	154
Calabria_laboratories tests Catalogues – LOINC 2.54 Mapping	2016	-	1,071
AIC – ATC Mapping	2017	-	18,309
Total		+ 571,500	475,970

2) The Integration layer: it is used for the phase of design and modelling of the data transformation process, semantic enrichment by means of external endpoints, and for the internal organization of data in the Knowledge Base. To this aim, we used the Kettle component and ETL procedures. In STI, ETL is a key process to bring all the data derived from code systems, which are heterogeneous in their structures and formats, in a standard, homogeneous environment. In particular, during the transformation phase, imported code systems are manipulated to be compatible with the target system (CTS2 model). In some cases, the necessary transformation rules are trivial, but in other situations (as happened for AIC and LOINC files, which changed database structures over successive releases) it may be necessary to sort, unite, and aggregate data. Pentaho Kettle provides a wizard that guides in the migration process, defining the source database server, the destination one, the mapping of the data types, and so forth, so that migration does not cause data loss.

The population of the Knowledge Base can be:

- manual, through the Web Application interface, by means of the compilation of specific forms and the selection of the relationships and classifications in the Knowledge Base;
- through Rest services. The system will allow to import resources within the Knowledge Base;
- semi-automatic, through the use of specific ETL systems that guarantee the information extraction and enrichment by querying external services, and finally adding the data to modelled knowledge.

3) The Semantic layer: it is based on the use of ontologies to extend data related to the atomic units (concepts in STI) with external components that have the related knowledge (e.g., the ontologies related to LOINC or ICD-9-CM concepts available in Bioportal [33]). To this aim, the Virtuoso platform was used. In particular, for each LOINC, ATC, or ICD-9-CM concept, it is possibile to query the Virtuoso platform in order to retrieve additional semantic/ontological information (e.g., the semantic type class of the LOINC concept *Hemoglobin A*, code “45208-6” is *Amino Acid, Peptide, or Protein*, or for example the LOINC concept *Aciclovir*, code “1-8”, is a pharmaceutical substance whose semantic type is *Nucleic Acid, Nucleoside, or Nucleotide*).

4) The Presentation layer: it is the interface that uses the Knowledge Base, characterized by the conceptual entities and enriched by a series of information. In particular, each concept in the Knowledge Base was enriched by the following information provided by using different panels in the interface:

- code details, including information derived from the structure of the code system;
- status and versioning, including information on the changes of the status of a code and on the different versions available in the system;
- relationships, including the relationships of a precise concept to other concepts in the code systems (e.g., the hierarchical relationships of the three digit ICD-9-CM code 282 *Hereditary hemolytic anemias* to its leaf codes, etc.);
- mapping, where all the mappings of the selected code/concept to other resources available in the STI Knowledge Base are visualized, if present;
- HL7 Specifications, including information needed to exchange data according to HL7 standard, i.e., Code, Code System OID, Code System Name, Code System Version, and concept Display Name;
- ontology, which gives access to the components LodView, to visualize the RDF data of a concept, and LodLive, to navigate the graph of a concept in the ontology derived from the external resource Bioportal.

An important aspect of the STI Knowledge Base is that all the resources, where applicable, were imported in bilingual versions. In particular, LOINC, ATC and ICD-9-CM are available in English and in their official Italian version. The Italian version is in most cases aligned to the corresponding English one. Exceptions are LOINC, where the Italian translation, as all other international LOINC translations, is always aligned to the previous LOINC English version; and the Italian translation of ATC, available in the system as version 2014, since it was not possible to collect the last Italian updates by the responsible government agencies.

C. *Web services development*

Considering the CTS2 functionalities described in Section II.A, we selected and implemented the following services in STI:

- *Reading*: reads the list of resources in STI and shows complete information on a single resource;
- *Search*: allows to search the resources for keywords or in particular fields thanks to the application of personalized filters;
- *Import*: allows to import the resources (being them code systems, value sets or mappings) into the Knowledge Base; the dataset within the CKAN component; RDF/OWL graphs into the Virtuoso triple store;
- *Export*: allows to export complete resources or the results of specific queries in CSV or JSON formats;
- *Update*: allows the editing of the Knowledge Base content;
- *Mapping*: allows the visualization of the existing mappings or the editing of new cross-mappings between the resources in STI;
- *Editorial workflow*: allows the approval of a particular resource or change in the Knowledge Base, as well as the validation of new mappings created by users with special permissions. The

validation can be carried out only by administrative users.

The listed services are to be used through specific Rest Services useful for the reuse of the STI functionalities and for the interoperability with other systems. They will allow interoperability among the systems Liferay-CKAN-Virtuoso. For the development of these services, the Spring Web MVC framework was used. It provides Model-View-Controller (MVC) architecture and ready components that can be used to develop flexible and loosely coupled Web applications. The MVC pattern results in separating the different aspects of the application (input logic, business logic, and UI logic), while providing a loose coupling between these elements.

In a Linked Data perspective, STI service allows semantic enrichment of a resource, thus obtaining relationships with related resources, by querying exposed SPARQL endpoint, such as the Bioportal one.

D. *Web Application Development*

Regarding the Web Application, the system covers the following functionalities:

- User registration, authentication, roles and permissions management.
- CMS management (platform browsing management, content and versioning management, etc.).
- Multilingualism/bilingualism management (possibility to switch from Italian to English language when browsing a resource).
- Resource utilization by means of the Reading Web service.
- Search of one or more resources or concepts by means of the Search Web service.
- Resource and workflow management.
- Import and Export.
- Mapping between resources.
- Use of ontological resources and graphs visualization by means of Lodlive.
- Browsing of the RDF resources with LodView, linked to the Virtuoso SPARQL endpoint.
- Use of SPARQL endpoint for the resources stored in Virtuoso. Virtuoso will expose the resources in the RDF/TTL format via the SPARQL endpoint, allowing the most technical and expert users to make more sophisticated queries.
- Management of the STI dataset imported in CKAN.

The Web Application allows to navigate the available resources according to the type (i.e., Code System, Value Set, Mapping). After the selection of a specific resource, depending on the original structure, it is possible to navigate the hierarchical tree, to directly select a code and visualize its details; or to search on the selected code system by using filters or full text search. It has to be noticed that when looking for a specific LOINC code, the service returns not only the exact matched code but also all the LOINC codes that have been redirected to it for guiding the users' mapping choices. This is particularly useful as it shows at the same time and for the same concept deprecated or discouraged terms and their correspondent new reference code, facilitating also double check and updates of the already performed mappings. The interface of the navigation functionality was built taking inspiration from the cited *Terminology Server*, the CTS2 implementation provided by the abovementioned University of Applied Sciences and Arts of Dortmund.

V. Results and Evaluation

STI service was released in its beta version in April 2017, as both Web service and Web application. It contains four standard code systems, which are those prescribed by the Law Decree regarding FSE, in their Italian and English versions, and also some mapping resources (local mapping to LOINC and AIC-ATC mappings). They can be accessed through the CTS2 main functionalities, such as searching, querying, navigation. Versions available are both those fixed by the cited law (i.e., LOINC 2.34 – December 2010) and the most recent ones (i.e., LOINC 2.63 – December 2017), so users can choose which one best fits their needs.

The service is open to the possibility of uploading additional code systems, mapping and value sets. They will be integrated, as it was for the four standards already available in the STI, taking into account their peculiar structure so to ensure a proper use of them. Furthermore, more local files mapped to the standard code systems can be uploaded by the system administrators after validation of their correctness. Regarding the mapping, there is also the chance, for users with special permissions (e.g., physicians, laboratory technicians, etc.), to create mappings between the available resources directly through the STI platform by using the *Cross-Mapping* functionality. During the cross-mapping, users have to qualify the mapping that they are creating between two concepts belonging to two different code systems, by selecting the type of association between the two selected concepts (e.g., choosing if two concepts are synonyms, clinically correlated, or if one is the hypernym of the other, etc.). These cross-mappings, in any case, will be validated by the system administrators before becoming effective and saved in the STI Knowledge Base. Regarding the interoperability services, as said in Section IV, STI allows external applications, e.g., other terminology services installed at a regional level, to make requests to the Web service, which are those provided by the CTS2. In particular, the following examples are given:

1. **Entity Description Query Service**

Example: Search the entity "*Immunoglobulina"* in the code systems ICD-9-CM and LOINC:

- http://sti.iit.cnr.it/cts2framework/entities?matchvalue=immunoglobulina&page=0&maxtoreturn=20&codesystem=ICD9-CM
- http://sti.iit.cnr.it/cts2framework/entities?matchvalue=immunoglobulina&page=0&maxtoreturn=20&codesystem=LOINC&format=json

Parameters:
* matchvalue= a string for fulltext search or a query in the Lucene syntax for more complex search on indexed fields
* page= page number (starting from 0)
* maxtoreturn= number of elements per page
* codesystem= code system to query (mandatory)
* codesystemversion= code system version to query (optional)
* format= required format (e.g., "json")

2. Code System Version

Example: LOINC version 2.56:

- http://sti.iit.cnr.it/cts2framework/codesystem/LOINC/version/2.56/entities?matchvalue=immunoglobulina&page=0&maxtoreturn=20&format=json

3. Entity Description Read Service

Example: Read the detailed information of AIC code 19227038:

- http://sti.iit.cnr.it/cts2framework/codesystem/AIC/version/16.01.2017/entity/AIC:19227038

4. Association Query Service

Example 1: Existing cross-mapping associated to the ATC v. 2014 code "B02AA01".

- http://sti.iit.cnr.it/cts2framework/associations?list=true&codesystemversion=2014&sourceortargetentity=B02AA01&format=json

Example 2: List of mappings between the code systems LOINC v. 2.54 and ATC v. 2014

- http://sti.iit.cnr.it/cts2framework/associations?list=true&changesetcontext=LOINC (2.54) - ATC (2014)&format=json

5. Entity Description Query Service

Example: List of mappings between a local code system (e.g., Umbria) and LOINC v. 2.54:

- http://sti.iit.cnr.it/cts2framework/codesystem/LOINC/version/2.54/entities?page=0&maxtoreturn=250&matchvalue=LOCAL_CODE_LIST:Umbria&format=json

6. Export Service

Example: Export in JSON format the Italian version of LOINC v. 2.58 and of AIC v. January 2017:

- http://sti.iit.cnr.it/cts2framework/exporter?codesystem=LOINC:2.58&language=it&format=json
- http://sti.iit.cnr.it/cts2framework/exporter?codesystem=AIC:16.01.2017&aictype=classe_h

The developed terminology service also has a FAQ section including the most common questions about code systems used in the FSE setting, questions about the Service and CTS2 itself, and finally, a Documentation section where it is possible to consult or download technical documentation about the service, such as the User Manual, documents related to the STI design and development, and a tutorial explaining its use and functionalities.

In order to test the functionalities and suitability of STI, we recruited a sample of test users, belonging to some of the Italian Regions that already implemented the FSE infrastructure. On one hand, we provided special permissions to Domain Experts (e.g., General Practitioners and Laboratory technicians) in order to let them use both free functionalities (e.g., concept search, navigation of the resources, download) and the Cross-Mapping functionality, to create clinical/semantic mappings directly through STI. On the other hand, we asked regional technical referent users to query the Web service from their local application to make requests such as those provided above (e.g., to have the list of all the *map_to* codes in order to verify if some of their mappings changed the LOINC reference code). Figure 3 and Figure 4 show respectively the search and cross-mapping activities performed by a domain expert (more specifically a Laboratory technician) for the concept *Glucose.*

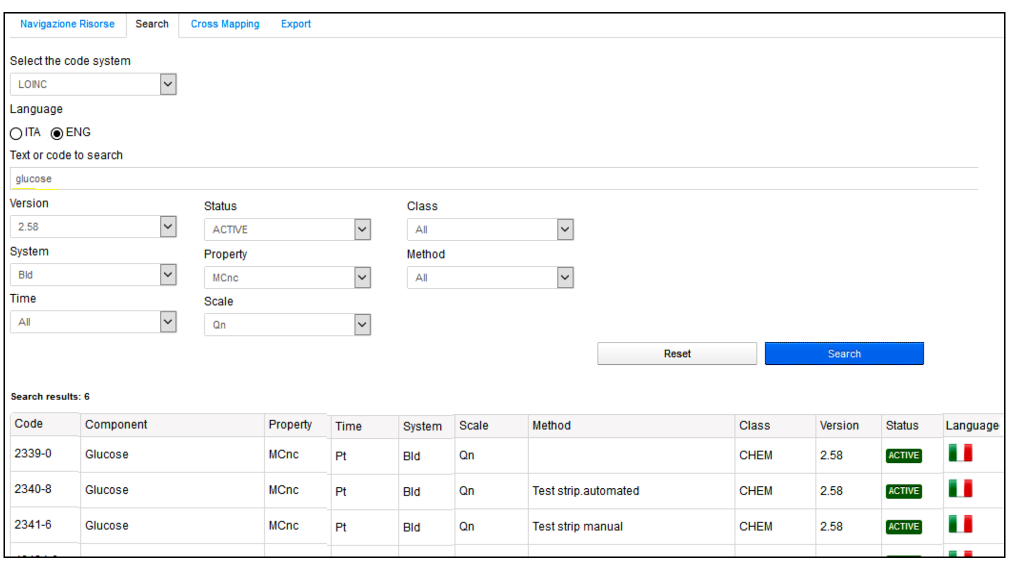


Figure 3. STI Web Application screenshots showing the Search functionality for the concept "Glucose" in LOINC

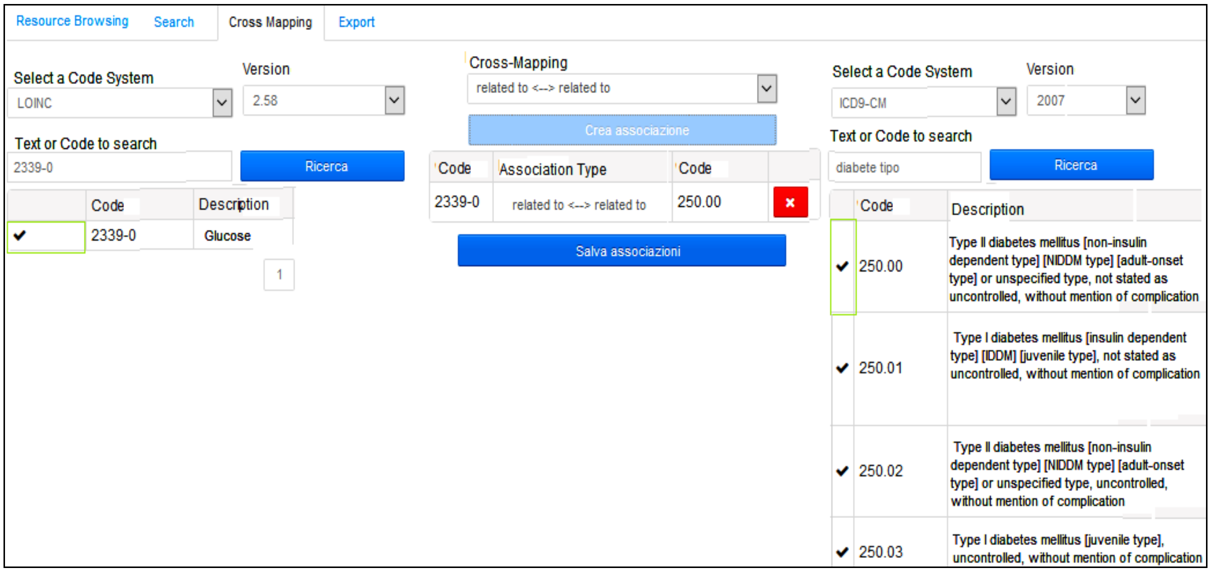


Figure 4. STI Web Application screenshots showing the Cross-mapping between the LOINC code 2339-0 "Glucose" and ICD-9-CM

In a Linked Data perspective, the STI users with appropriate permissions are able to manage graphs and, in the future, will be able to import data in RDF format within the STI Virtuoso platform. To support Virtuoso, OpenRefine is proposed [34], for guaranteeing the connection to ontology models, and thus the connection of the CSV fields in the STI DBMS to the corresponding RDF. Figure 5 shows a scheme that clarifies this workflow. This will permit to access and reuse STI resource in additional and innovative ways:

- through the browsing of RDF resources using LodView, which will be connected to the SPARQL endpoint of STI Virtuoso platform;
- through LodLive, allowing the user to navigate and explore a graph within the Linked Data Cloud, as shown in Figure 6.

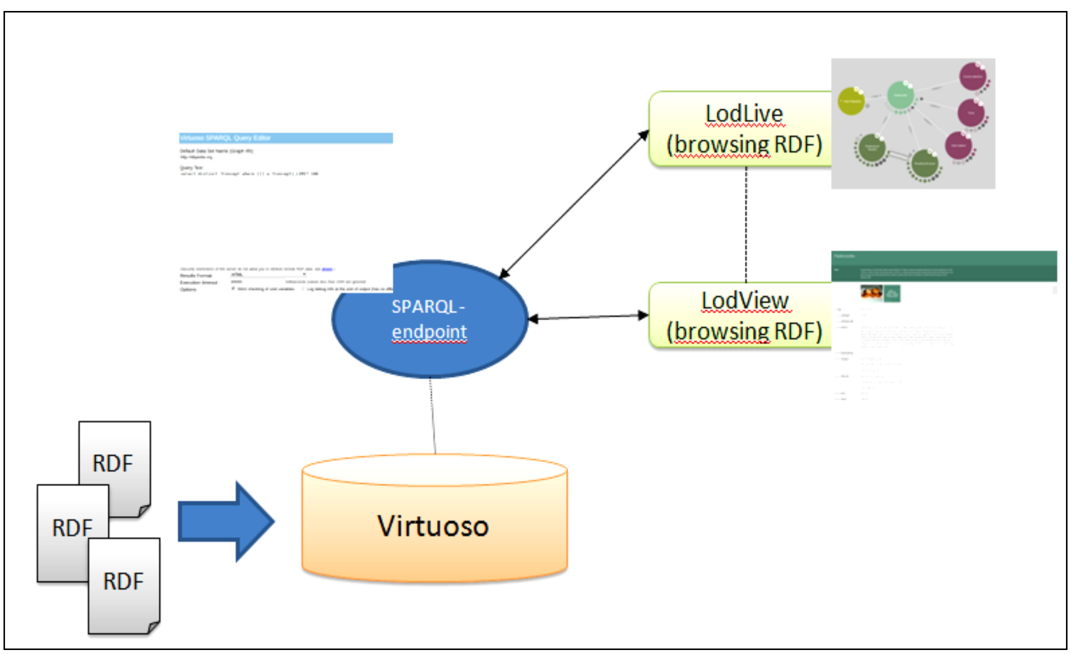


Figure 5. Data import scheme within the STI Virtuoso platform and graphs browsing

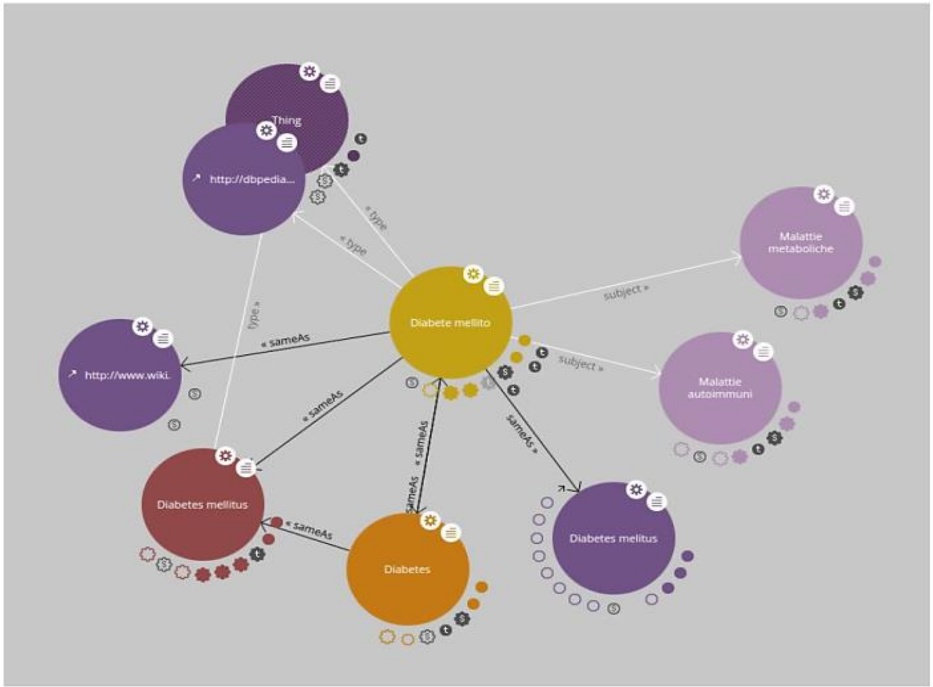


Figure 6. Navigation of the graph related to the concept “Diabete Mellito” in the resources exposed by STI Virtuoso.

VI. DISCUSSION AND CONCLUSIONS

This paper describes the design and development of a bilingual (Italian – English) integrated terminology service, named STI, based on the CTS2 HL7 standard. The service includes for now the four code systems required by the FSE Law Decree, but it is open to the possibility to integrate further terminologies in the future.

Designing a terminology service is a non-trivial pursuit, especially when resources with different structures need to be integrated and available for different uses. This was the first issue of this work, as it required a personalized design and implementation for each code system uploaded into the STI Knowledge Base. For example, LOINC has multiple informative axes, which were reported into both the main visualization screen (the six fundamental axes) and an openable window tagged with different labels. Nonetheless, importing LOINC into the service was challenging because its database structure changes as versions evolve. So, a preliminary normalization step was carried out in order to uniform names and values of the fields of the different versions.

Moreover, when dealing with AIC, as the system is released in four separate files, ETL procedures needed to be trained to import each of them every time there is an update, and check if mappings to ATC are present in the new AIC files or if they need to be extracted by following the procedure described in Section IV.B.2). All the above mentioned issues are an obstacle to the flexibility and scalability of the service. Furthermore, it was not always easy finding updated versions of the four code systems, especially in computable format, such as CSV files, and for some of them both master English and translated Italian files are not available (i.e., ATC). The chance to visualize ontology representation of the clinical terminologies is not usable for all the versions of the systems. This is an interesting possibility offered by the STI that needs to be improved in the future releases of the service. Efficiency and effectiveness of an EHR also depend on the possibility of unambiguously exchanging and understanding incoming information.

Semantic interoperability improves significantly thanks to the implementation of a terminology service, especially if it is compliant to a standard such as HL7 CTS2, which is widely adopted. The services offered (e.g., searching, querying, and cross mapping) are particularly useful when national or local code systems need to be linked to standard classification systems. This interoperability also strongly depends on the alignment between terminologies and their quality. This work shows the path that has been taken, also thanks to the recent advancements promoted by the law and by the AgID and CNR collaboration within the context of FSE projects, to align Italian FSE with international initiatives that promote the use of integrated management services of medical terminologies.

Nonetheless, it has to be considered that the implementation of integrated terminology services is just the beginning of the process. In fact, the most important aspect in managing medical terminologies is the maintenance over time to update resources and coordinate processes such as transcoding, translation, and licensing.

In fact, maintenance of such a system is the real challenge: systems change, errors are made, and the lifecycle of mappings and data must be considered. Sometimes, mappings can be contextual and absolute consistency is very hard to achieve. That evidences the need for a dedicated governmental authority that coordinates the entire process.

Among the several advantages provided by STI use in the Italian FSE framework there are:

- the possibility to share official terminologies, and their updates between the FSE central node and the services used by the local/regional FSE nodes;
- the possibility to configure policies (roles and authorizations) and to model the organization of the system (concerning production/editing of the terminological resources) through terminology management roles;
- the compliance of the data model and application services with the CTS2 standard (Normative Standard CTS2 Version 1.2);
- the delivering services of terminological resources with standard protocols and formats (JSON, CSV);
- the possibility to make advanced searches with personalized filters according to the code system selected, and to find additional semantic information by navigating their ontological graphs;
- the distribution as open source tool, with a GNU GPL license.

Some of these advantages and functionalities characterize STI if compared to existing CTS2 implementations, especially at a national level. In fact, the terminology services cited in Section II.A, even if more sophisticated from a technical and architectural point of view (e.g., in the cited DiTAM service, the possibility to have many local terminology service nodes connected to the central DiTAM node in a federated network), are: proprietary, thus more difficult to be used by a Public Administration; less precise in the structuring and visualization of the code systems; and, to our knowledge, do not allow the access to the resources as Linked Data, or their ontological graphs; and, finally, they do not provide bilingual access as provided in STI.

Among the improvements currently underway on the STI, we can mention: i) an extension of the service data model, and of the Knowledge Base, in order to guarantee modeling, semantic integration and managing also of other types of code systems used at the national level but also local code systems (regional laboratory catalogues, regional catalogue for prescriptions and services, etc.); ii) the definition of a general structure for importing and mapping in order to make the service more flexible and scalable; iii) the definition of new ETL processes rules for the code systems integration and mapping; iv) the development of new functionalities supporting the maintenance of the system like the CTS2 *History* (i.e., the ability to determine which changes occurred over stated periods of time) and the *Update* services (the ability to validate load sets of changes into the service that updates its content); v) the improvement of the cross-mapping functionality to allow mapping editing by applying new association types; vi) the extension of the export functionality also to query results and mapping tables (currently the service allows exporting available code systems, in their different versions, but does not allow to export extracts of them or query results tables).

The ability to share, query and maintain official and up-to-date terminological artifacts using an accepted standard terminology service interface, such as STI will allow standard terminology content to be readily disseminated and validated, and becomes more useful as organizations (healthcare facilities, Regions, Ministry of Health, and national Standard Development Organizations) in the FSE context begin to undertake the enhancement and maintenance of terminologies to support language translations, jurisdictional extensions to standard code systems, or maintenance and development of local terminologies, avoiding the proliferation of heterogeneous resources, and local tools and technologies to manage terminologies. The use of STI as an open source Terminology Service in the FSE context, whose functioning is directly linked to the level of interoperability and the degree of security of the sensitive data processed among the different Regional systems, gives the chance to Regions and local healthcare facilities to comply with a series of regulations and, in particular, with the modifications to art. 68 of the Legislative Decree 82/2005 - "Digital Administration Code" (CAD) [35], all aimed at giving a preferential road to the use of free software.

Finally, the creation of a terminology management service, such as STI, to be used in the context of the Italian FSE, is not only a way to reach semantic interoperability, but it represents a better support to healthcare professionals for improving the quality of clinical data ensuring maximum benefits along the healthcare process and the cooperation among different healthcare providers.

ACKNOWLEDGMENT

This work is supported by the projects *Realizzazione di Servizi della Infrastruttura Nazionale per l'Interoperabilità per il Fascicolo Sanitario Elettronico* (prot. IIT 0005723, 01/09/2015) and *Realizzazione dei servizi e strumenti a favore delle Pubbliche Amministrazioni per l'attuazione del Fascicolo Sanitario Elettronico* (agid.AOO-AgID.REGISTRO UFFICIALE (I).0021033.31-12-2015) funded by the Agency for Digital Italy. A special thanks goes to LINK Management and Technology S.p.A. for the support in the development of the STI Web Service.

REFERENCES

[1] E. Cardillo and M. T. Chiaravalloti, "Towards the Development of a CTS2-based Terminology Service in the Italian Federated Electronic Health Record", SEMAPRO

2017 : The Eleventh International Conference on Advances in Semantic Processing, 2017, pp. 12-21.

[2] M. Ciampi, A. Esposito, R. Guarasci, and G. De Pietro, “Towards Interoperability of EHR Systems: The Case of Italy”, ICT4AgeingWell, 2016, pp. 133-138.

[3] E. Cardillo, M. T. Chiaravalloti, and E. Pasceri, “Healthcare Terminology Management and Integration in Italy: Where we are and What we need for Semantic Interoperability”, European Journal of Biomedical Informatics, vol. 12 (1): pp. en84-en89, 2016.

[4] Prime Ministerial Decree [Law (general)] n. 178, 29 September 2015, “Regolamento in materia di fascicolo sanitario elettronico. (15G00192)”, GU Serie Generale n.263, 11-11-2015.URL: http://www.gazzettaufficiale.it/eli/id/2015/11/11/15G00192/sg [accessed: 05-15-2018].

[5] ANSI/HL7 V3 CTS R2-2015, HL7 Common Terminology Services - Service Functional Model Specification, Release 2, February 2015, URL: https://hssp.wikispaces.com/specs-cts2 [accessed: 05-15-2018].

[6] European Community, “Directive 2011/24/EU of the European Parliament and of the Council of 9th March 2011 on the application of patients' rights in cross-border healthcare” [Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:088:0045:0065:EN:PDF].

[7] Semantic Interoperability for Health Network FP7-ICT-2011-7, URL: http://www.semantichealthnet.eu/ [05-15-2018].

[8] C. Daniel, D. Ouagne, E. Sadou, K. Forsberg, M. Mc Gilchrist, E. Zapletal, et al., “Cross border semantic interoperability for clinical research: the EHR4CR semantic resources and services”, AMIA Jt Summits Transl Sci Proc. 2016, pp. 51–59.

[9] Trillium Bridge II - Reinforcing the Bridges and Scaling up EU/US Cooperation on Patient Summary, H2020-EU.3.1.5. - Methods and data, URL: http://www.trilliumbridge.eu/ [accessed: 05-15-2018].

[10] CTS2 Development Framework, Mayo Clinic Informatics, URL: https://github.com/cts2/cts2-framework [accessed: 05-15-2018].

[11] Fast Healthcare Interoperability Resources (FHIR), Health Level 7 Specification, URL: https://www.hl7.org/fhir/ [accessed: 05-15-2018].

[12] O. Bodenreider, “The Unified Medical Language System (UMLS): integrating biomedical terminology”, Nucleic Acids Res: vol. 1; 32(Database issue); pp. D267–D270, 2004 DOI: 10.1093/nar/gkh061.

[13] J. Grosjean, G. Kerdelhué, T. Merabti, and S. J. Darmoni, “The HeTOP: indexing Health resources in a multi-terminology/ontology and cross-lingual world”, 13th EAHIL Conference, 2012, ceur-workshop, vol. 952, 2012. URL: http://ceur-ws.org/Vol-952/paper_17.pdf [accessed: 05-15-2018].

[14] J. Pathak, H. R. Solbrig, J. D. Buntrock, T. M. Johnson, and C. G. Chute, “LexGrid: A Framework for Representing, Storing, and Querying Biomedical Terminologies from Simple to Sublime”, Journal of the American Medical Informatics Association: JAMIA. vol. 16(3), pp. 305–315, 2009, DOI:10.1197/jamia.M3006.

[15] J. B. Lamy, A. Venot, and C. Duclos, “PyMedTermino: an open-source generic API for advanced terminology services”, Studies in Health Technology and Informatics, 2015, vol. 210, pp. 924-928.

[16] D. Luna, G. Lopez, C. Otero, A. Mauro, C. Torres Casanelli, and F. González Bernaldo de Quirós, “Implementation of Interinstitutional and Transnational Remote Terminology Services”, AMIA Annu Symp Proc, 2010; pp. 482–486.

[17] D. K. Sharma, H. R. Solbrig, E. Prud’hommeaux, K. Lee, J. Pathak, and G. Jiang, “D2Refine: A Platform for Clinical Research Study Data Element Harmonization and Standardization”, AMIA Jt Summits Transl Sci Proc. 2017; 2017, pp. 259–267.

[18] L. Zhou, H. Goldberg, D. Pabbathi, A. Wright, D.S. Goldman, C. Van Putten, et al., “Terminology modeling for an enterprise laboratory orders catalog”, AMIA Annu Symp Proc., 2009, pp. 735-739.

[19] K. J. Peterson, G. Jiang, S.M. Brue, and H. Liu, “Leveraging Terminology Services f or Extract-Transform-Load Processes: A User-Centered Approach”, AMIA Annu Symp Proc. 2017, pp. 1010-1019.

[20] Standard Terminology Services (STS), PHAST, URL: http://www.phast-association.fr/accueil_doc_sts/ [accessed: 05-15-2018].

[21] C. Seerainer and S. W. Sabutsch, “eHealth Terminology Management in Austria”, Studies in health technology and informatics, vol.228, pp. 426-30, 2016.

[22] Terminology Server, University of Applied Sciences and Arts, Dortmund, URL: http://www.wiki.mi.fh-dortmund.de/cts2/index.php?title=Hauptseite [accessed: 05-15-2018].

[23] DiTAM, Codices S.r.l.. 2014. URL: http://www.codices.com/prodotti/ditam.html. [accessed: 05-15-2018].

[24] R. Gazzarata, M. E. Monteverde, E.Vio, C. Saccavini, L. Gubian, I. Borgo, and M. Giacomini, “A Terminology Service compliant to CTS2 to manage semantics within the regional HIE”, EJBI – Vol. 13(1), pp. 43–50, 2017.

[25] International Classification of Diseases, 9th Revision Clinical Modification (ICD-9-CM), Italian version, Italian Ministry of Health URL: http://www.salute.gov.it/portale/documentazione/p6_2_2_1.jsp?lingua=italiano&id=2251 [accessed: 05-15-2018].

[26] Anatomical Chemical Therapeutical (ATC), WHO Collaborating Centre for Drug Statistics Methodology Norwegian Institute of Public Health URL: https://www.whocc.no/atc_ddd_index/ [accessed: 05-15-2018].

[27] Autorizzazione all’Immissione in Commercio (AIC), Agenzia Italiana del Farmaco. URL: http://www.agenziafarmaco.gov.it/en . [accessed: 05-15-2018].

[28] Logical Observation Identifiers Names and Codes (LOINC), Regenstrief Institute, URL: https://loinc.org/ [accessed: 05-15-2018].

[29] Liferay 6.2., URL: https://www.liferay.com/downloads . [accessed: 05-15-2018].

[30] Pentaho Data Integration (ETL) a.k.a Kettle, URL: https://github.com/pentaho/pentaho-kettle [accessed: 05-15-2018].

[31] Virtuoso, OpenLink Software, URL: https://virtuoso.openlinksw.com/ . [accessed: 05-15-2018]

[32] CKAN, URL: https://ckan.org/ . [accessed: 05-15-2018].

[33] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, et al., “BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications”, Nucleic Acids Res. 2011 Jul;39, pp. W541-5. Epub 2011 Jun 14. URL: https://bioportal.bioontology.org/ [accessed: 05-15-2018].

[34] Open Refine, URL: http://openrefine.org/ . [accessed: 05-15-2018].

[35] Legislative Decree 7 March (2005), n. 82. Codice dell'amministrazione digitale. In GU Serie Generale n.112, 16-05-2005 - Suppl. Ordinario n. 93 [available from: http://www.agid.gov.it/sites/default/files/leggi_decreti_direttive/dl-7-marzo-2005-82_0.pdf].

Supporting Collaborative Care of Elderly through a Reward System based on Distributed Ledger Technologies

Emilien Bai and Kåre Synnes
Department of Computer Science, Electrical and Space Engineering
Luleå University of Technology
Luleå, Sweden
e-mail: emibai-6@student.ltu.se, unicorn@ltu.se

***Abstract*—This paper discusses supporting collaborative care of elderly through a reward system based on distributed ledger technologies. The design and implementation of such a reward system that connect elderly and volunteers by mutual agreements involve technologies such as smart contracts and blockchains. The work is motivated by the demographic change, where an aging population consequently increases the need for care. This causes a great tension in our society, as care resources become increasingly constrained, both regarding costs and availability of care staff. Much of the daily care of the elderly is today done by family members (spouses, children) and friends, often on a voluntarily basis, which adds to the tension. The core idea of this work is to help broaden the involvement of people in caring for our elderly, enabled by a system for collaborative care. The proposed system benefits from recent advances in distributed ledger technologies, which similarly to digital currencies, are build on the ability for mutual agreements between people who do not know each other. The system also benefits from recent gamification techniques to motivate people to collaborate on a larger scale through performing simple daily tasks. The proposed system benefits from inherent distributed ledger technologies advantages, such as a high level of decentralization, thus a high availability, and strong data consistency. These advantages make it interesting to develop the possible links between blockchains and the outside world to allow for a higher level of automation and distribution of services such as collaborative care. New models for distributed ledger technologies, such as Iota tangles or the Swirld platform, may however scale and perform better than blockchains. These should thus be considered for a full implementation and test of the system. In summary, this paper presents a novel framework and prototype implementation of a reward system supporting collaborative care of elderly, that is based on distributed ledger technologies.**

Keywords-component; Blockchain; Collaborative Care; Gamification.

I. INTRODUCTION

The work in this paper in based on a paper presented at the UBICOMM 2017 conference [1]. The aging population has been identified as a challenge for the future in a Swedish study from 2013 [2]. In May 2012, 18.8% of the total population of Sweden was 65 years old or older. This part of the population is expected to reach 20.5% in 2020 and 25.9% in 2060. The main difficulties identified are to finance welfare of the aging population, as well as meeting the increasing demand of service provision. The demand on staff is expected to increase by 210 000 caregivers by 2030 in Sweden, while the supply is expected to stay quite the same. Also, this situation will probably result in a widened financial gap between the cost of welfare and state revenues. The trend of an aging population is confirmed to be worldwide by a United Nations report from 2015, which focuses on the oldest persons (aged 80 years or more) [3].

Much of the daily care (such as performing daily tasks like shopping for groceries, cleaning, cooking, etc) are often performed by informal carers such as family members, or friends. The burden this places on spouses and children of the elderly can often be very high, reducing the quality of life not only for the elderly being cared for but also for these informal carers.

It is thus clear that a broader engagement of our society in caring for our elderly is needed, where voluntary contributions also can be rewarded (besides the altruistic satisfaction of being helpful, pro-bono). Not everyone would of course require such rewards, but motivating a larger cohort of our fellow people may require both short and long term perceived benefits. Examples of short term benefits may be making people's contributions visible in the society or being able to trade work, and long term benefits may include being able to get help back in kind (If I help now, then I will get help later). This leads to the following research question:

How can a system for collaborative care of elderly be designed and implemented to engage and motivate people to contribute with daily tasks on a voluntary basis?

The aim of this work is thus to develop an application intended to connect the population who may need help in common daily tasks with people who may provide voluntarily help. The aim is not to replace workers specialized in health care, but to reduce their work charge where it is possible and therefore instead leave them more time to do important and skilled tasks for the elderly.

Ultimately, by reducing the proportion of paid care, the application may also contribute to decreasing the cost of care for the aging population, without degrading the quality of care.

The rest of the paper is organized as follows. In Section II, we present a state of the art concerning distributed ledger technologies (DLT), blockchains as well as smart contracts. In Section III, we introduce the methodology used in order to develop the system. Section IV focuses on the implementation and design of the system. In Section V, we present the design of the gamification aspect. In Section VI, we discuss how the designed system fills the needs of our research question and point out some limitations. Finally, we conclude this paper in Section VII.

II. State-of-the-art

The rapid digitization of our society is key to alleviating the tension on our care systems, where recent technological and methodological advances bring great potentials to enable an increasingly collaborative care. One example is communication technologies, where access to mobile computing now is nearly ubiquitous and where we now at any time can engage in our social networks. A recent example is DLT, including the notions of Blockchains and Smart Contracts, which are introduced in this section together with novel methodologies for user engagement, namely Gamification.

A. Distributed Ledger Technologies

A distributed ledger (also called a shared ledger) is a consensus of replicated, shared and synchronized digital data geographically spread across multiple nodes (sites, countries or institutions) [4]. There is no central administrator or centralized data storage. Instead, a peer-to-peer network is required together with consensus algorithms to ensure that replication and consistency is maintained across the nodes of the distributed ledger.

The most popular distributed ledgers are based on public or private blockchains, which employ a chain of blocks to provide secure and valid achievement of distributed consensus. The first Blockchain was conceptualized in 2008 by Satochi Nakamoto [5] and implemented in 2009 for the digital currency Bitcoin. The example of bitcoin demonstrates the huge potential of blockchains for mutual agreements between two parties without the need of a trusted third party. For example, the volume of daily bitcoin transactions has been over 175 000 since April 2016 [6]. However, the bitcoin blockchain only scratches the surface of the potential of the technology, as it is focused and dedicated on the exchange of value, in the form of bitcoin transactions.

Distributed ledger technologies are expected to have a disruptive effect in our society, especially concerning mutual agreements, as they show many advantages: 1) agreements made on top of the blockchain do not need a trusted third party, and 2) each transaction needs to be signed by its sender using asymmetric encryption, which removes the need of an authentication layer in applications as this is directly handled at the blockchain level. It could ease the exchange of property between people or allow a more fine-grained digital right management.

B. Blockchains

Blockchains are distributed databases for transaction processing, and they are well suited for financial transactions but not limited to such applications. The use of blockchain technology also extends to non-financial applications and is for example, considered for supply chains, asset management or electronic health records.

All transactions are stored in a single ledger and ordered by time. The ledger represents the current state of the system and is replicated across every node. The transactions are broadcasted to the network and accepted if valid, by distributed consensus mechanisms, and are then grouped into a block, which is to be added to the blockchain. The last, and key, operation is to compute an ID for this block before storing it on the blockchain.

This operation can be done by solving a mathematical problem (usually random with a low probability), based on the previous block index (this takes around 10 minutes for the Bitcoin blockchain). The problem consists in finding a nonce (an integer value) to associate with the hash of the content of the block and the id of the preceding block. Once these 3 values concatenated, the resulting hash of this concatenation must respect a constraint: being less than x, x evolving in order to keep a relatively constant period between each block. This constraint can only be fulfilled by trying new solutions for the nonce. Once an ID has been computed, the network adopts the block and begins to work on finding the ID of the next block. The process of computing an ID in this way is called **proof-of-work** and it makes the blockchain immutable since changing an existing block requires to compute the ID for all the following blocks while the blockchain continues to grow. One of the weakness in the proof-of-work mechanism is the 51% attack. In the case an organization controls more than 50% of the computing power, it can start censuring transactions an can refute mining outside of the organization, as the blockchain considered as valid is the one replicated on the majority of blocks, in order to centralize all the rewards.

This operation can also be done using a **proof-of-stake**, where the miner is chosen in a deterministic way. One of the proof of stake design, used in Peercoin [7], is based on the concept of “coin-age”. The coin age is a number, which depends on the product of coins times the duration they have been held by the node. The higher the coin age is, the bigger the chances to be elected for the associated node are. Once a node has been selected to mine a block, the duration it held the coins is reset in order to avoid the richest and oldest nodes from dominating the blockchain. This method is more energy efficient than a proof-of-work as it does not imply any power competition between the nodes. It is also safer against attacks as acquiring the majority of the coins is usually more costly than collecting 51% of the computing power of the network.

As the technology evolves, new consensus mechanisms appear. We can cite Proof-of-Elapsed-Time (PoET) [8], used in Hyperledger Sawtooth [9]. This consensus mechanism is based on trusted function called at the central processing unit (CPU) level. It reproduces a leader election protocol found in many consensus mechanisms and distributes leadership across the population of validators. This mechanism has a low cost of implication, which increases the potential number of validators and therefore the robustness of this consensus algorithm.

Every transaction on a blockchain is signed, using asymmetric key cryptography, ensuring its provenance. The nodes in the network check the transaction conformity (an user can only spend the money he owns from previous transactions and can only perform a transaction in his own name: this is verified thanks to the cryptographic signature), authorization (an user can only perform a transaction in his own name) and resistance to censorship.

If a transaction conforms to the protocol, it will be added to the ledger without any party being able to discard it. All transactions can be processed peer-to-peer without the need of a trusted third party, since a blockchain network does not rely on any central authority but on a distributed consensus.

C. *Public and Private Blockchains*

This study began with a review of existing blockchain technologies, which revealed two main categories of ledgers: public or private.

A public blockchain is a ledger for which anyone executes transactions or mines blocks. Since anyone can modify a public blockchain, they offer a high replication rate. This is also what makes public blockchains slow and less energy-efficient. Since anybody can contribute to public blockchains, it also offers pseudonymity where an user is only identified by an address and all the transactions referring to this address can be read. Some projects are being developed in order to create a true anonymity when using a Blockchain technology. For example, that is the goal of Zcash blockchain [10] that protect the privacy of its users using zero-knowledge privacy.

A private blockchain does not allow everyone to join. It usually belongs to a single company, or a group of companies, running the chain and validating transactions. It usually uses a certificate authority in order to control the access and the rights of each stakeholder. The level of decentralization is not as good as in public blockchains, but performance is generally significantly higher. Indeed, when located on a public blockchain, a decentralized application represents only a small proportion of the entire system instead of representing the majority of it and, therefore, gains efficiency. It allows for greater privacy since users are chosen and known. However, private blockchains are therefore not resistant to censorship. The main entity running the blockchain can decide to stop one stakeholder to execute transaction.

D. *Permissioned and Non-permissioned Blockchains*

Our study also identified two subcategories of blockchains: permissioned or non-permissioned. In a permissioned blockchain, each node has a limited role. It may only be allowed to validate transactions, mine new blocks, execute smart contracts (see below) on the blockchain or perform transactions with the chain assets. On the contrary, a non-permissioned blockchain allows any node to take any role. Table I illustrates the resulting categorization of studied blockchains.

TABLE I. CHAIN CLASSIFICATION

	Non-permissioned	**Permissioned**
Public	Ethereum[11], Bitcoin[5], Iroha[12]	Ripple[13]
Private		Fabric[14], Burrow[15], Openchain[16], Multichain[17]

E. *Smart Contracts*

In 1994, Nick Szabo defined smart contract as [18]:

"A smart contract is a computerized transaction protocol that executes the terms of a contract. The general objectives are to satisfy common contractual conditions (such as payment terms, liens, confidentiality, and even enforcement), minimize exceptions both malicious and accidental, and minimize the need for trusted intermediaries. Related economic goals include lowering fraud loss, arbitrations and enforcement costs, and other transaction costs."

Smart contracts are a way to enforce a legal agreement without the need of a trusted third party. It consists of computer code, stored on the blockchain and its execution can change the state of the blockchain. A sample of smart contract code demonstrating a simple use case of an asset holder contract is illustrated in Figure 1. It profits from blockchain immutability to ensure that the terms cannot be modified. Thus, smart contracts cannot be modified and the result of an interaction is predictable and not corruptible. A high level in data integrity (as well as a good log level in case of breach in contracts design) is therefore ensured. Smart contracts develop the need of experts able to formalize legal agreements and convert it in clear and complete specifications. These specifications have to be translated in computer code and audited in order to ensure that all corner cases are covered.

```
pragma solidity ^0.4.11;
contract Bank {
    // We want an owner that is allowed to selfdestruct.
    address owner;
    mapping (address => uint) balances;
    // Constructor
    function Bank() public {
        owner = msg.sender;
    }
    // This will take the value of the transaction and add to the senders account.
    function deposit() public payable {
        balances[msg.sender] += msg.value;
    }
    // Attempt to withdraw the given 'amount' of Ether from the account.
    function withdraw(uint amount) public payable {
        // Skip if someone tries to withdraw 0 or if they don't have enough
        // Ether to make the withdrawal.
        if (balances[msg.sender] < amount || amount == 0)
            return;
        balances[msg.sender] -= amount;
        msg.sender.transfer(amount);
    }
    function remove() public {
        if (msg.sender == owner){
            selfdestruct(owner);
        }
    }
}
```

Figure 1: Sample of smart contract code

F. Bitcoin and Etherum Smart Contracts

The Bitcoin blockchain only runs a single smart contract, where the Bitcoin blockchain only ensures that the sender actually owns the tokens he wants to send in a transaction. As a consequence, a receiver cannot refuse a transaction. It is usually not referred to as a smart contract since the code is more embedded inside the chain protocol than actually running on a chain virtual machine (VM).

Ethereum is the biggest public platform for smart contracts [11] which provide a Turing complete language for smart contracts executing in a virtual machine environment. As it is a public and uncensored platform, all users are free to send their own code, to be executed by the Ethereum VM. To avoid malicious user from locking the system, executing infinite loops algorithm for example, the VM use a “gas” system. When sending code to be run by the VM, users have to send a certain amount of gas associated with it. For every computation cycle required by the contract execution, a small amount of gas is consumed. If there is no more gas to consume, the computation is stopped, and an error is returned. Gas is bought using the chain currency (in our case ether): Its consumption is used to reward the nodes who took part in the smart contract execution. Smart contracts are triggered by receiving a transaction and they process transaction data in order to change the state of the contract. The code execution is replicated on each mining node in order to validate the transaction and include it in the next block. Code execution happens as many times as there are nodes validating transactions. Therefore, there are some limitations in the smart contracts design [19]: it is difficult to link smart contracts with outside world events, or make use of external services automatically (without a transaction). If a contract is waiting for data from the outside world and it does not receive the same data on every node, this would create a conflict on the chain. That is why smart contract are mainly triggered by transactions that ensure data consistency across the nodes. However, one library, Oraclize is aimed at bridging external service with smart contract code in a secure way, but has not been tested in the field of this study. On the contrary, if a contract must call an external Application Programming Interface (API), which will trigger an action, it cannot determine which node is responsible to actually make the call. This design issue can be avoided using external logs watcher that can check a contract status then trigger the outside chain action if needed.

Smart contracts are thus not well suited to ensure agreements outside the chain, but they remain a very efficient way to condition fund transfer inside a chain. We can imagine an internal coin system with which users agree on a value and use it to limit the volume of “official” currency. This mirrors the current financial system where major stakeholders like banks agree on the value of a debt toward each other and balance the debt volume without actually exchanging assets. Finally, smart contract are also not well suited to hide confidential data, especially on a public chain. Every node replicates the database, and can try to brute-force encryption of data if need be. Also, every transaction is relatively anonym, which means activity of user towards contracts can be traced.

G. Involver

Our technical review only identified one mobile application for volunteering, an application called “Involver”. It is defined as a social volunteering platform [20].

The goal of this application is to bring together potential volunteers with partner organizations that need help. Every cause a volunteer can help with is ordered based on location, subjects and skills needed. The rewards are brought by sponsors and take the form of non-monetary advantages.

The application also offers to certify the number of volunteering hours on professional social networks. It also includes a social aspect emphasizing the fact that volunteering is more interesting with friends. This example illustrates the need of a trusted third party (in this case the application) when agreements are made between volunteers and organizations.

Involver is however more of a start-up than a scientific platform. It is also not aimed at manipulating sensitive data, as this kind of information is to be held out of the application, by the organizations themselves if need be.

III. METHODOLOGY

This work is based on multiple theories within the gamification field, which form the basis of the theoretical work as well as the implementation. This section describes the definitions in the context of this work.

A. Gamification Definitions

The word gamification appeared for the first time around 2002, when Nick Pelling used it for its consultancy business [21]. Gamification is according to Hutoari and Hamari [22]*”a process of enhancing a service with affordances for gameful experiences in order to support user’s overall value creation.”*

Deterding et al. [23] propose a more general definition of gamification as ”*the use of game design elements in non-game contexts”*. This definition is supported by the distinction made between games and play, with gaming being more structured by rules and more competitive. Game elements are defined as elements that are characteristics to games, found in most (but not necessarily all) games and found to play a significant role in gameplay.

Since gamification has been a trending topic, it prompted a lot of academic studies, which showed gamification to be present in many different contexts such as learning (e.g., Duolingo [24]), exercise (e.g., Fitocracy [25]), work and more.

B. Gamification, Rewards and Volunteers

The gamification aspect is part in the final application as an incentive for volunteers to use it. A review of studies concerning gamification by Hamari, Koivisto, and Sarsa [26] show that, globally, gamification has positive effects and benefits on users where it is used. Gamification have a positive impact on the behavior of users, but also a psychological impact, acting on motivation, attitude or enjoyment of users while filling tasks. The gamification aspect is expected to motivate users and maintain their involvement.

Gamification can also be coupled with rewards in order to extend the scope of the gamification to the real world. A reward system can thus be seen as a natural part of a gamification system. For example, it can be used as a mechanism to engage, motivate and compensate users who volunteer their time and services for collective purposes.

Volunteers' motivations to help have been shown to be more based on intrinsic rewards (to fulfill psychological needs). However, small and non-expensive rewards are also appreciated and encouraging, where, generally, the goal is not to spend as much in a reward as the cost to pay someone [27]. The vision of combining gamification with real-world rewards allows reaching both fully altruistic people, who probably would not use the rewards or would not see it as an essential part of the application, as well as an audience needing more recognition to maintain its motivation Rewards are also expected to stimulate the interest of people usually that do not usually take part in volunteering activities.

C. Achievements and Badges

Achievements are a really common part in gamified applications. They are usually associated with badges and find their origin in merit-badges given to boy scouts of America since 1911. In 2003, Wikipedia started Wikipedia's Barnstars [28], aimed at rewarding contributors for their involvement on the platform. Another example of successful use of achievements is the Foursquare badges [29], which encouraged people to complete tasks in real life in order to unlock them. Furthermore, all games published on the Microsoft Xbox Live [30] platform are required to have achievements. A study by Anderson et al. [31] showed that badge placement in an application can have an effective influence on user behavior and also affect his/her use of the application. However, a study by Montola et al. [32] concludes that achievements globally have a positive effect on motivation but can sometimes be confusing for some users if they are not introduced properly. In summary, badges can be efficient incentives and are relatively cheap to implement in an application.

D. The Hamari and Eranti Achievement Framework

According to the framework designed by Hamari and Eranti [33], an achievement can be divided in three main parts.

Firstly, an achievement has a signifier, which is the visible part of an achievement and conveys information about it. It consists of a name that set the theme of the achievement and hints at the completion logic for it. The signifier also includes a visual, which completes the name and often has two states, unlocked where the visual is faded and completed where the visual gets fully colored. Finally, the signifier has a description, which describes what is required from the user to complete the achievement and what can be gained by completing it.

Secondly, an achievement also consists of a completion logic. It consists of a trigger, a pre-requirement (specific date, already completed achievement), a conditional requirement to determine if the action is triggered and also a multiplier, which determines how many times the three first parts have to be completed to unlock the achievement.

Thirdly, achievements carry rewards to show the user the achievement that has been completed. When added to a game, achievements completion can be a way to unlock in-game rewards. The external part of the reward is often the fact that these achievements are displayed publicly.

E. Leveling

Leveling based on experience points is an easy way for users to keep track of their progress. It was originally used in role playing games, and then extended to any type of games. The logic behind leveling is quite simple: when performing a task, users receive points, and then when a certain amount is reached, the user advances a level. In games, earning levels is often linked to gain or progress skills for the avatar. In a gamified application, advancing a level is recognition of the skills acquired by the user in real life: it can also allow an user to access more advanced features. In games, points are earned when completing a mission/quest: in gamified applications, points are delivered when the user completes the task the app is trying to help with. For example, points can be delivered when a volunteer completes an offer, based on the number of tokens earned. However, when the user spends his tokens, he keeps the same number of points.

An important part of a levelling system is the threshold: it represents the number of points needed to reach the next level. Usually, the first levels have a low threshold, in order to keep the user motivated and show quick progress. Then, once users have been significantly engaged, thresholds get bigger to be more challenging and therefore more rewarding.

IV. Implementation

Based on the pros and cons listed above, the Burrow blockchain has been chosen to conduct our test implementation. It is fast, provides a smart contract virtual machine and the permission layer allows controlling the access rights. As the system works with sensitive data, it benefits from the privacy a private blockchain provides. The permission layer allows limiting the number of nodes allowed to mine blocks or create contracts on the chain. Our chain is therefore only dedicated to our system and does not spoil resources for other contracts. It also uses a proof-of-stake consensus mechanism, which is better suited to the use of a private blockchain, since every mining node is known and trustworthy. A proof-of-stake consensus is also more energy efficient and faster than using proof-of-work.

A. System Goals

The designed system is intended to connect elderly with volunteers who can help them with everyday life tasks, which do not require any specialized skills (for example, in health care). The goal is not to replace health care workers but to reduce their workload where it is possible in order to give them more time for specialized tasks. As an incentive, volunteers receive a non-monetary reward, a token based on time spent to help and eventually resources involved in tasks, as the use of a vehicle for example. These tokens can be used to acquire rewards as coupons or advantages / discounts in local shops or non-monetary advantages. On top of this, gamification aspect using points and badges is added to keep volunteers motivated.

Users of the platform need to register by giving their identity. Their personal data is stored encrypted and only revealed to other users if they share a task (tasks can also be referred to as offers or task offers from elderly users). Once users are registered, an authority is charged of verifying the information and grants the permissions according to the user status. For example, this external authority has to check that volunteers who claim to have a driving license actually have it, and more generally verify the identity of users who register. This authority could follow the model used by the car sharing application for example. An elderly user is allowed to create task offers: they describe the mission, specify a time slot when it has to be performed and duration, and a type for the task (gardening, shopping, accompanying for visits, etc.). The reward amount is computed according to the offer specifications. A task that requires the volunteer to own and use a vehicle will be rewarded with a higher reward then task with no material need. Once the offer is created, volunteer users can see it and read its specifications: if a volunteer is available and able to fulfill the task, he or she can commit to it. From there, the elderly user can access the volunteer's contact information to schedule the task more precisely. Once the task is accomplished, the volunteer needs to claim the reward. The elderly user can then confirm that the offer has been fulfilled: this action triggers the issuing of tokens for the volunteer who helped. With these tokens, the volunteer will be able to buy rewards. Rewards are added by rewarder users. These rewards contain a description, a price and a code, delivered only when the reward is bought. This is illustrated in the activity diagram below, see Figure 2.

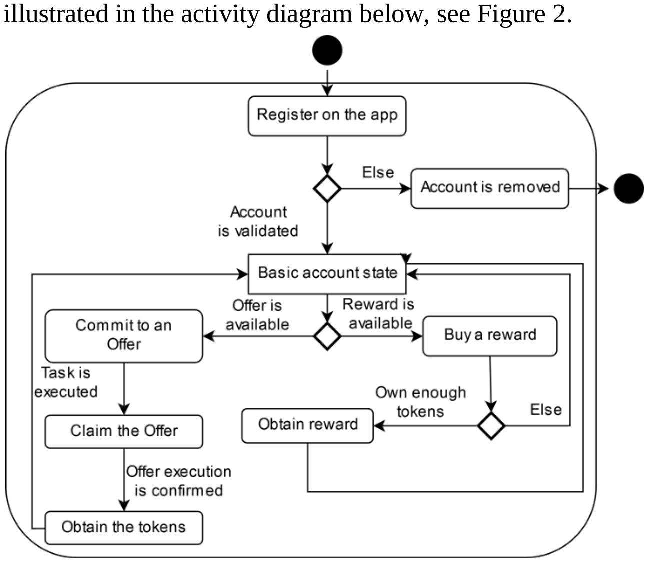


Figure 2: Activity Diagram for Volunteer Users

B. Global Design

The system back-end is built on top of a blockchain with smart contracts to handle agreements between the users. It has four main contracts handling the different parts of the system: these contracts form a database while they also ensure system consistency. These database contracts do not directly store data. They are more data structures that reference other contracts where the data actually is.

The bank contract handles the tokens for each user: the only way tokens are issued is when an elderly user confirms that a task offer has been fulfilled. The only way to use these tokens is when a volunteer user spends them to buy a reward. The bank contract stores the balance for each user and ensures that the user actually owns enough of them before spending them.

The user contract is used to store user data. One part of this data is readable by everyone (and uses pseudonimity) while sensitive data stays encrypted and is only revealed when a task links two users. This contract is also used to handle permissions for each user. Permissions are set by an authority according to the status of the user: depending of his permission level, an user can or cannot perform some actions in the application (As an example, only a volunteer user can commit and claim an offer).

A contract is used to store offers and commit, claim and confirm their execution. An offer is a task, proposed by an elderly user for which help from a volunteer is needed. Thus, offers are smart contracts with properties and states: the state of the offer evolves during the course of the agreements but properties are immutable. This evolution is described in Figure 3.

Finally, a contract is used to store and buy available rewards. Rewards are added by partner rewarders in a limited availability and bought by volunteers.

The blockchain handles authentication of users for these actions, allowing a mutual agreement between different users without the need of a trusted third party, once the registration is complete. The blockchain also guarantees the content of agreements since contracts cannot be discreetly modified. The division of the application is described in Figure 4.

C. Detailed Architecture

The system is built on the Monax blockchain, Burrow [15], a fork of the Ethereum blockchain allowing working with a permissioned ledger: this permission layer also allows using a proof-of-stake mining mechanism. Another difference compared to unpermissioned ledger is that nodes can have restrictions on how they can contribute: some can be dedicated to validate nodes, while others handle permissions or receive transactions.

Contracts are developed using Solidity [34], an object-oriented programming language for smart contract development. Solidity code of a contract needs to be compiled outside of the blockchain, and then is sent using a specific type of transaction. The result of this operation, if successful, is the address of the contract. This address will then be used to interact with the contract, by calling its functions in transactions or when reading its state.

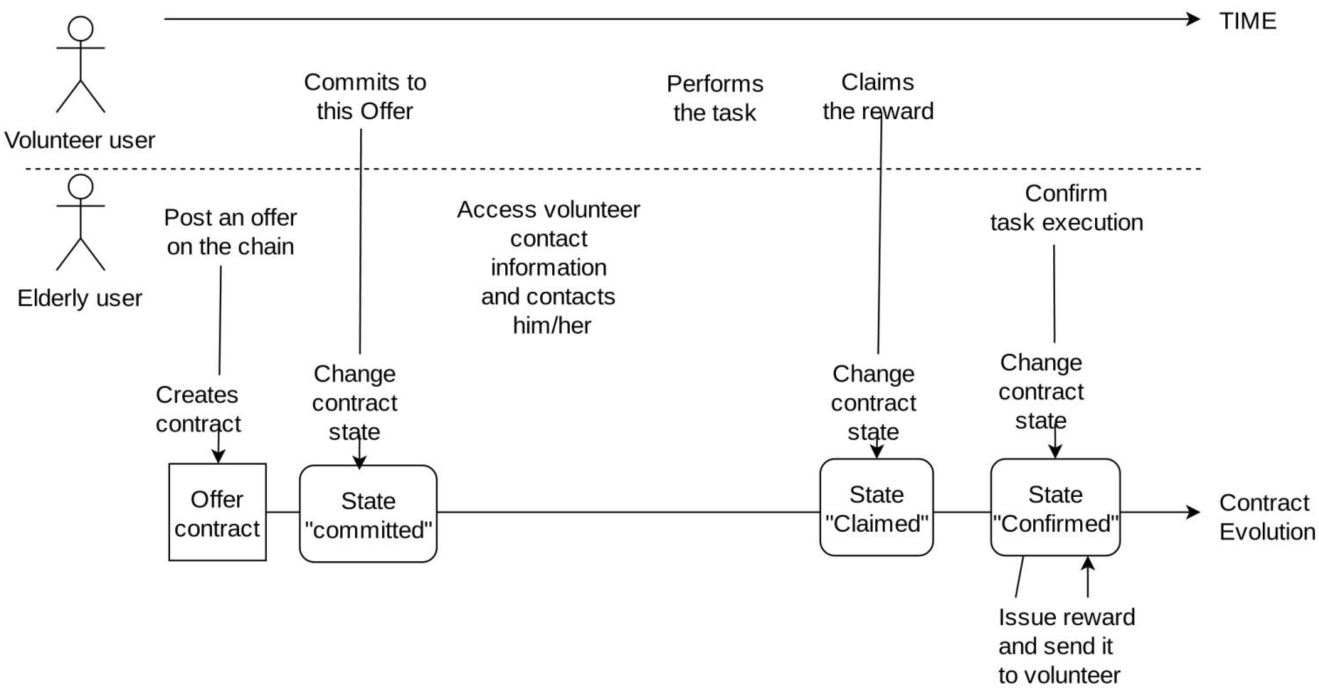


Figure 3: Offer evolution

The application is developed following an action-driven architecture coupled with a five types model.

The five types model suggests splitting the application using different kind of contracts. Database contracts, where data is stored, can be read or updated. As explained before, these databases are key value store for our data. They consist in lists where a contract address can be found using its id. The use of a custom data structure (our lists) over the one usually used when developing in Solidity (mappings that are dictionary already existing in the basic structures of Solidity) has been done in order to be able to query all the entries in a database, when we would have need to keep every data contract names to retrieve them if using the mapping structure. Controllers contracts operate on database contracts, and can operate on multiple databases (for example, read user's permissions from one database and then operate an action on another). A third type of contract is contracts managing contracts (CMCs) where other contracts addresses are kept in view and can be replaced if needed (if we update the code of a controller for example). They provide single point of entry to the system, which is useful when a system uses many contracts and therefore also many addresses. They include an update mechanism for controller contracts in order to be able to edit a code that would otherwise be immutable. These CMCs allow the update of existing controller contract in a transparent way for the user. Without them, when updating a controller contract, the user would need to obtain the address of the updated version of the contract. Using this system, the user only need to query the contract wanted using its name and the CMC directly redirect the query to the latest version. Application logic contracts (ALC) are contracts specific to an application and they perform multiple operations using controllers and other contracts. Finally, utility contracts can be seen as libraries: they perform a specific task, without modifying the state of other contracts and can be used without any restrictions. They are not project specific and can be re-used in different situations.

The five types model effectively separate actions used to interact with databases contracts. Actions are thus focused on small parts and modifications of the system. Actions are smart contracts with only one function (in our case, "execute") and perform atomic modifications to the system. It can be seen as a microservice architecture even though it does not share the goals of such an architecture but appears more as a need to be able to maintain and update the application. Actions are stored in a CMC. This architecture allows the system to be updated more easily than if using full controller contracts. A full controller would handle every interaction with database contracts. Any simple modification to a simple function implies the full contract replacement, which infers heavy interaction with the chain. As a result, actions can be dynamically replaced without the need of modifying a complete controller contract. As we can see in Figure 4, the update mechanism is integrated directly in the action driven architecture. Users with the right permission level can add, replace or remove an action from the system in the same way more general users interact with it. The main CMC is a Decentralized Organization Upgrade Guy (DOUG) storing all databases' contracts of the system and especially, the action database. The action database is also a CMC and works with an action manager calling this database in order to find the actions to execute. By following this architecture, an user can interact with the system knowing only the address of the DOUG and the databases public APIs. On the developer side, maintenance operations are

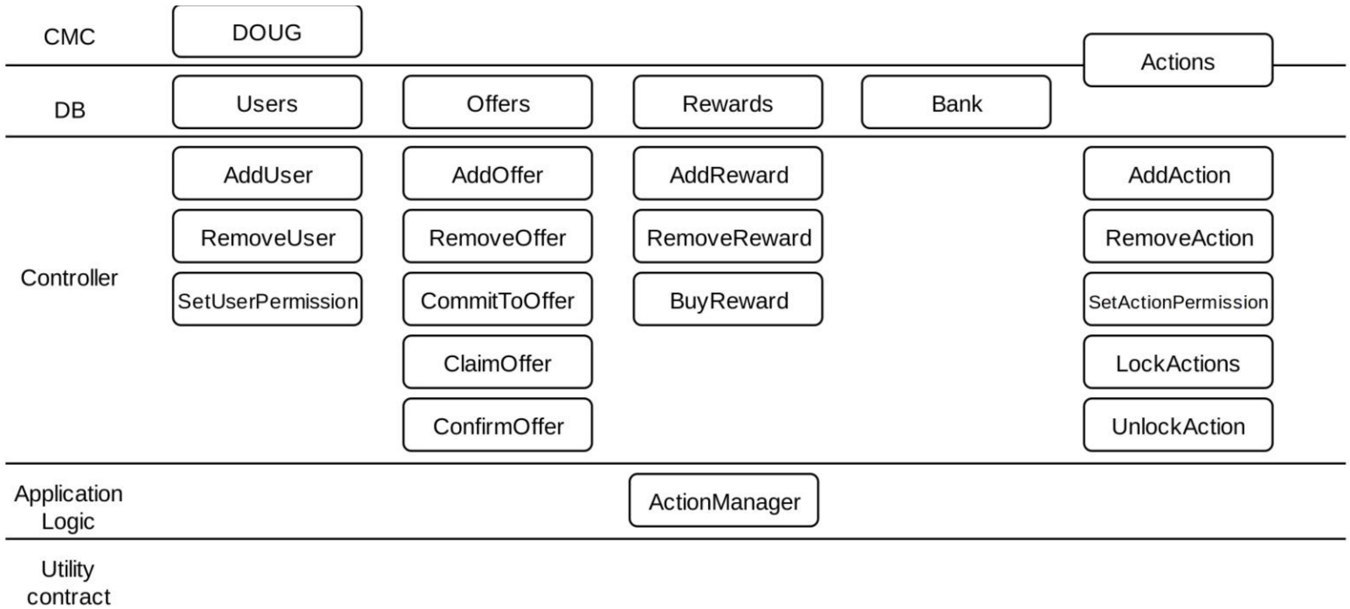


Figure 4: The Five Types Model

simple because actions can be easily updated (replacing an action in the database is an action itself) without the need of updating every interface as long as the DOUG contract stays the same (and therefore keeps the same address). Since action replacement is an integrated part of the application, this also allows using multiple developer accounts with the right permission level instead of locking the system by limiting updates to only its creator. The process of executing an action is described in Figure 5. First, users, that only know the address of the DOUG, query this contract in order to get the address of the action manager (1). Then, users then send a transaction to the action manager address using as parameters the action they want to execute (2). The action manager contract receives the transaction. It queries the DOUG in order to order to get the address of the action database (3). The action manager then queries the address of the action contract that will be executed (4). The action manager repeats step 3 to get the address of the user database (5), and repeat action 4 in order to get the calling user data and especially permission level of the user (6). From the data, it can read in the action contract, the action manager verifies that the caller user has the right permission level to execute this action (7). Then, it calls the execute function of the action contract if allowed to do so (8). The action contracts query the DOUG to get the addresses it needs to perform its action (9), and modifies the database data accordingly (10). Databases are locked in such a way that they can only be modified by the action being currently executed. The action returns the result of its execution to the action manager (11) that returns it as a result of the transaction originally sent by the user (12).

D. *Interface*

In order to keep the system as decentralized as possible, the user credentials are not kept in a database. The most suitable solution is to let users manage their own credentials and this can be done using a mobile or desktop application acting as a Bitcoin wallet. This solution presents as an inconvenient a high risk of credential loss. It should only be coupled with an efficient "save-and-restore" system that would allow users to keep a safe copy of their own. Another solution to store credentials can be paper wallets: these are cardboard-cards with flash-codes or text-written credentials. This can be a solution to effectively handle permissions and verify user information by sending them their credentials using for example, standard post, or asking them to present themselves to an office where the identity verification occurs and their paper wallet is delivered. The credentials on paper-wallets can afterwards be stored in a mobile application or required to be scanned for every action performed through the application. This solution has not been chosen in the prototyping step, but should be considered in a following step.

The user then interacts with the blockchain through a Representational State Transfer (REST) API. Every node used as a validator for the blockchain is also used to host an API server, allowing a good level of decentralization. The choice of using a REST API comes from technical limitation at the moment the project has been conducted. An ideal design applies the creation, signing and sending of the transactions directly from the device it is sent, and not centrally executed on a third part server. The use of an installed application instead of a web based application also limits the number of request needed for developing the gamification aspects, since, as a Bitcoin wallet, this kind of application does not need to store every transaction but only those concerning users. The developed application used for testing is illustrated in Figure 6.

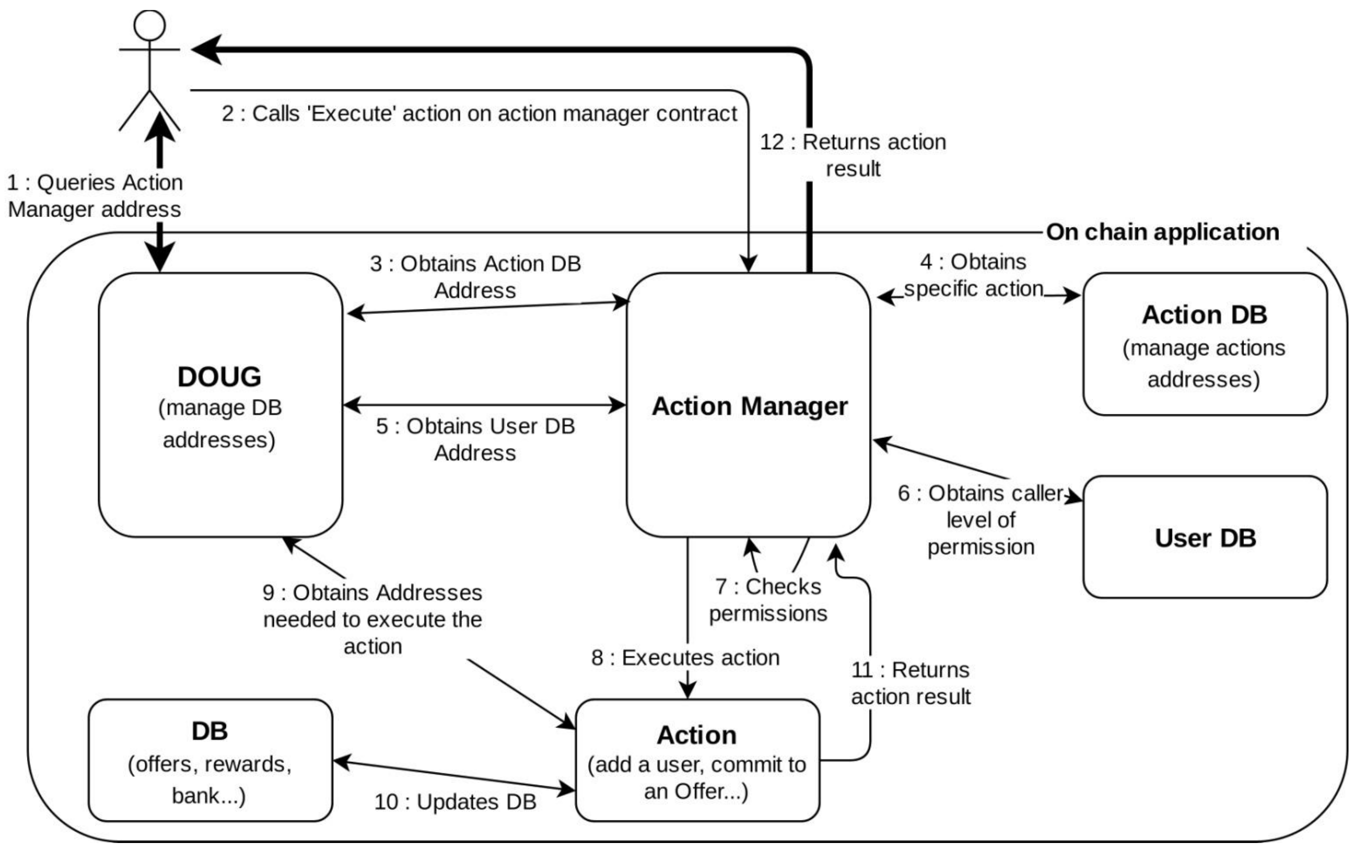


Figure 5: Execute Action Process

E. *Technical limitations*

The system was first designed to be as modular as possible: every action had its own execute method, taking various numbers of arguments. Since this modularity implies no real inheritance, the action manager had to use low-level calls, a Solidity feature where a function is called at a specified address without knowing the contract API at the caller level. These low-level calls return a boolean only indicating if the call succeeded (a function has been found and called) or not, but not the actual return value. This solution has been abandoned since it created a lot of problems in data formatting for arguments at the action manager level and afterwards at the action level. The absence of a relevant return value was also a performance issue since it implied to check every action execution afterward. Finally, the choice has been made to use a formatted schema for the execute method of every action, covering all the current cases, and ignoring some useless parameters for some of the actions contracts. This choice reduces genericity but improve the reliability and performances of the system.

Some gamification aspects have been limited by the use of smart contracts as databases. This layout is not really efficient when querying many contracts and therefore limits some features such as ranking between all users. This limitation is also linked to our choice of using list as our main data structure. This choice is easy to deploy and reliable but does not scale really well or not really well suited to filter elements.

V. Gamification Design

The choice for our application is to limit the gamification aspect to the frontend in order to reduce the volume of interaction with the back-end and therefore improve performance. It has been a design choice from the beginning, as there were no existing studies concerning performance and scalabity of decentralized applications, either on public or private blockchains. Badges fit really well with this vision. Achievements are well oriented towards volunteers, as they are the target we try to motivate. The achievements implemented have two main objectives. The first objective is to serve as a tutorial, where these achievements appears when doing really basic actions (such as to commit to an offer or getting a validated account) and are supposed to show the possibilities of the application while also introduce the achievement system. This kind of rewards is supposed to be numerous in the beginning: it guides the users toward using all the features of the application and rewards them quite often in order to provoke a feeling of significant progress and create engagement. The second objective is to maintain motivation and encourage involvement. The achievements for this objective are focused on quantity and regularity: it consists in fulfilling a defined number of offers, typed or not: buying rewards, keep checking the app and keep fulfilling offers every week/month to create strikes.

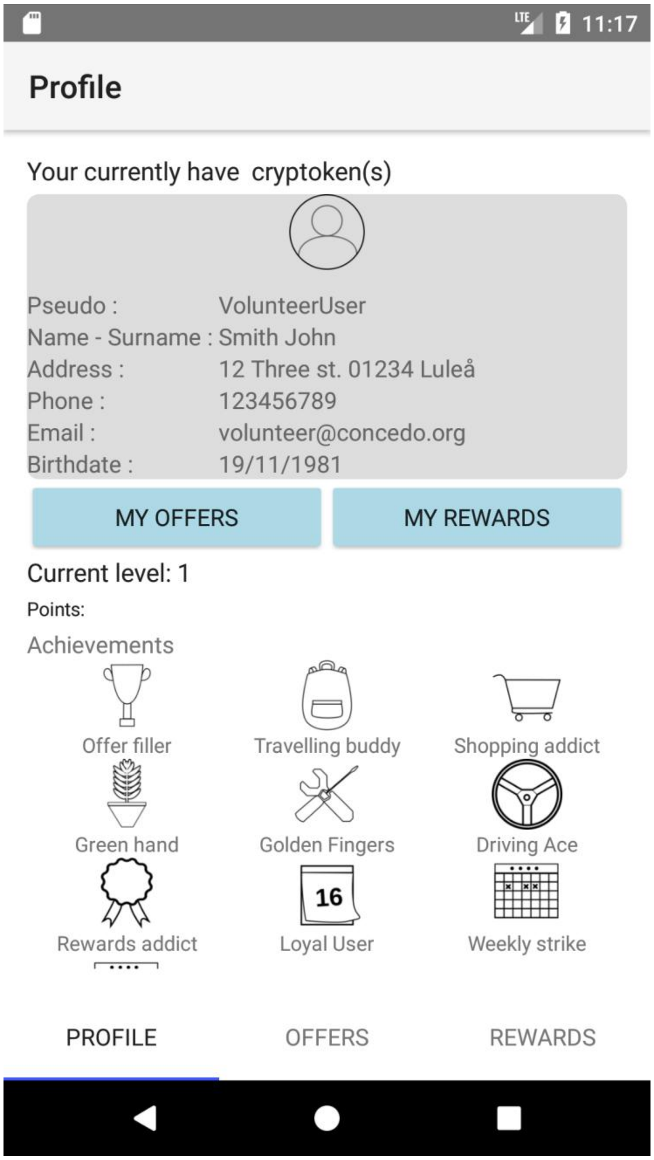


Figure 6: Application Interface for Volunteers

The achievements implemented in the application are detailed in Table II.

TABLE II. LIST OF ACHIEVEMENTS

Achievement / Multiplier	Step 1	Step 2	Step 3
Complete X offers	10	50	100
Buy X rewards	10	50	100
Complete X offers- Gardening	5	20	50
Complete X offers- Shopping	5	20	50
Complete X offers- Driving	5	20	50
Complete X offers- DIY	5	20	50
Complete X offers- Accompanying	5	20	50
Use the app for X days	30	180	360
Complete X offers in a week	2	4	6
Complete X offers in a month	4	8	15

The Hamari and Eranti Achievement Framework, as detailed above have been utilized to design the achievements for the application. Since no formal study have been found regarding leveling curve formulas, and since in game examples are often based on experience points gains varying depending on level, the following formula has been chosen in order to progressively increase the levelling thresholds:

$$t_1 = 30$$
$$t_{n+1} = t_n + \frac{l_n}{5} \quad (1)$$

The chosen initial threshold (to pass from level 1 to level 2) is 30. Since points are approximately equivalent to minutes, this allow users to gain levels quite quickly at first, then require more engagement to level-up further. For example, the users need a bit less than 4 hours of cumulated engagement to reach level 5. Reaching level 10 requires around 13 hours for the volunteer and reaching level 20 is almost 90 hours. Having such a progression is aimed at challenging users who will need to dedicate more and more time in order to level up. The progression of points needed is illustrated in Figure 7.

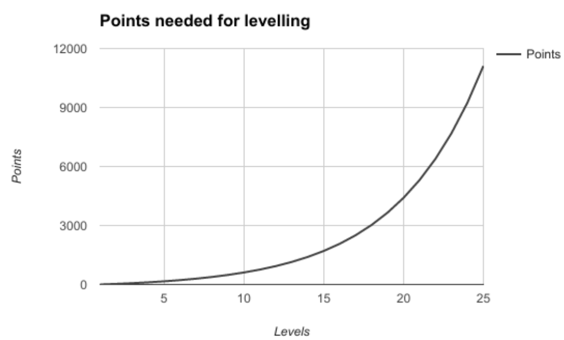


Figure 7: Levelling Curve

Levels can be seen as a simplified achievement since it can be associated with a name and a visual status. The chosen completion logic, focusing more on global progress, is however simplified in comparison to an achievement focused on specific task completion. Gaining points also happens more often than unlocking an achievement.

VI. DISCUSSION

How can a system for collaborative care of elderly be designed and implemented to engage and motivate people to contribute with daily tasks on a voluntary basis?

A system for collaborative care of elderly should come as a complement to the "classical" care system. It means that it should be efficient to help professional workers while keeping the costs low and the elderly population safe. The costs problem can be partially treated by implying volunteers in the process of elderly care and creating a system where moderation needs are low. In order to keep these cost low, the system needs to be reliable and able to work with a very high level of autonomy. DLT appear to be interesting by their performance in handling mutual agreements between parties not knowing each other. Such a system only requires moderation at registration and then can handle itself efficiently. Such an application should be mobile in order to integrate itself efficiently in daily life for both volunteers and rewarders. This mobile deployment is really compatible and almost comes as a complement for such a distributed system.

Even if the system benefits from the advantages of the blockchain on mutual agreements, it also shows some limitations due to the fact that it requires many interactions with the outside world. Firstly, the designed system can be abused by two parties knowing each other and agreeing on hypothetical tasks in order to issue tokens. This bias can partly be solved by limiting the number of offers an elderly user can post every week/month. The risk is to disadvantage honest users who need a lot of help, while only curbing abuses. However, since every commitment can be publicly visible, these abuses could be detected and user banned although this implies more authority regulation than just certifying user identity at registration. Therefore it reduces the interest in using such a trust-based system.

Another issue comes from the fact that the system does not allow nuances: a task offer will either be confirmed or not. Even if task confirmation could be coupled with a notation system, weighting the reward would be really dependent on personal appreciation. In the worst situation, an offer is not completed at all and this case is not automatically disadvantageous for the abuser while it can have negative consequences for the elderly user. Nevertheless, this situation is the same in every system implying trade between nonprofessional users (such as in carpooling services for example), and require an impartial arbiter to be resolved. However, what differs from another service is the criticality of the failure from one of the user. If a volunteer does not execute a task, it can have critical consequences for the elderly user while the volunteer only faces being banned or some equivalent penalty. It would require a legally recognized contract agreement in order to avoid this kind of situations. Such a protocol could discourage the potential volunteers. It could also create confusion between volunteers and professional workers.

Finally, another bias that could possibly appear is the preference for the most rewarding offers at the expense of the smaller ones. Even if rewards are calculated based on the efforts needed for their fulfillment, the least demanding offers could be discouraging because of the external efforts it can imply.

In this paper, we discuss many use cases where DLT could have a disruptive effect compared to our current applications. The main aspect is the focus on removing the need of a trusted intermediary in different kind of exchanges (monetary, intellectual property, assets management). However, we already have and use solutions daily to tackle this kind of scenario. And problems that come from theses scenario are usually not linked to a lack of trust in the intermediary. In our financial system, the vast majority of population trusts the banks and does not need any kind of censorship resistant money that would require users to engage more time and energy than what we currently have for quite a similar result. Moreover, the crypto-currencies based on blockchains as we design them today are not compatible with the functioning of the financial system based on debt and monetary creation. It would require either a huge adaptation from the financial system to fully embrace crypto currencies based on blockchain (which is highly unlikely) or a transformation in the paradigm we use to design crypto currencies and therefore, would result in the loss of what made a huge part of their interest: being a digital cash. Another interesting feature of smart contracts is the automation level they can provide. Yet, this level of automation is also available in our current systems when based on an external trusted authority. DLT could only enable the deletion of this intermediary but not actually creating new processes.

Not to be fully negative on the subject, we can foresee interesting use cases concerning bookkeeping and trace-ability. DLT may reveal themselves interesting when it comes to store important and non sensible data in an immutable way for an extended amount of time for example. Private blockchains could also be put in use to synchronize swarms of machine working together independently and not requiring a centralized synchronization. But even if the problems we approach today using DLT are “non-problems”, it is still an interesting field of research as we need time to explore the full capabilities of this relatively new technology.

VII. Conclusion

Care of elderly is an important and sensitive topic, which raises many and various concerns. The need for care will grow and volunteering will have to take part in this care in order to maintain reasonable costs for the society, as well as a sufficient level of services. A service to establish contact between people needing help and people willing to volunteer therefore is well motivated. This kind of system creates mutual agreements between users not necessarily knowing each other and can therefore take advantage of DLT.

Smart contracts are most efficient with on-chain agreements but show limitations when interfacing with outside-chain events. Interfacing with such events requires additional control points during the course of the agreements, to keep consensus between users. This reduces the interest in comparison with traditional, often centralized, systems between non-professional users.

In conclusion, the system described here still benefits from inherent DLT advantages, such as a high level of decentralization, thus a high availability, and strong data consistency. These advantages make it interesting to develop the possible links between blockchains and the outside world to allow for a higher level of automation and distribution of services such as collaborative care.

VIII. Future Work

This proof-of-concept system and prototype would be required to be evaluated with real users, first in small scale through participatory design and then in larger scale to ensure statistical certainty of results. The current project can be seen as a feasibility study to pave the way for this future work that would require more extensive resources.

The next development step in this project would consist in working on a new agreement protocol that could for example, involve a third-party user of the system to settle and confirm or not the task execution. This could be associated with an appreciation system working both ways: the elderly user evaluates the service he received in terms of motivation or punctuality (the goal is not to judge the skills) while volunteers could rate the offer description accuracy and the reception received. By adding a rating system for both tasks and users, it should motivate users to provide a quality service and filter abusive users or at least point them out. Using smart contracts, we can imagine to automatically suspending accounts who received a very bad appreciation in order to clarify the situation with the authority running the service.

We could also think of adding more filters to offers based on the elderly user preferences and needs: for example, some tasks may require a valid driving license that can be authenticated at registration: then, only users with a valid driving license would be able to see and commit to these offers. One last feature that can be investigated is a bidding system allowing volunteers to compete on committing to an offer and, therefore, not base the system on a first-come, first serve model. This would probably allow a selection focused more on the volunteer motivation.

A future step would also be to include more in-life elements, starting with paper wallets in order to authenticate users for the first meeting or while claiming a reward. Offer passwords, needed to claim the offer, shortly evoked in the precedent paragraph taking the form of matrix bar-code only readable by the mobile application could be used as a proof of the meeting while complicating frauds.

Finally, the concepts of utilizing IOTA tangles (iota.org) or the Swirld platform (swirlds.com) to replace blockchains as mechanisms to implement trust and data sharing needs to be further investigated. These could mitigate the inherent problems with many of the popular blockchains used today, such as scalability and performance issues.

IX. ACKNOWLEDGMENTS

This work was sponsored by Vinnova and supported by Ericsson Research in Luleå through the academy-industry exchange project Concedo. The work has previously been published in a shorter paper at UBICOMM 2017 [1].

REFERENCES

[1] E. Bai and K. Synnes, “A Reward System for Collaborative Care of Elderly based on Distributed Ledger Technologies”, In the proceedings of the Eleventh International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM), November 12-16, 2017, Barcelona, Spain , [ed] IARIA, ISSN: 2308-4278, ISBN: 978-1-61208-598-2.

[2] P. M. Office, “Future challenges for Sweden”, 2013, [Online], Available: http://www.regeringen.se/49b6cf/contentassets/389793d478de411fbc83d8f512cb5013/future-challenges-for-sweden--final-report-of-the-commission-on-the-future-of-sweden, [retrieved: 10, 2017].

[3] U. N. D. of Economic and S. A. P. Division, “World population ageing 2015”, 2015, [Online], Available: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf, [retrieved: 10, 2017].

[4] UK Government, Office for Science, “Distributed Ledger Technology: beyond blockchain”, 2016, [Online], Available: http://www.ameda.org.eg/files/gs-16-1-distributed-ledger-technology.pdf, [retrieved: 09, 2017].

[5] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system”, 2008, [Online], Available: http://bitcoin.org/bitcoin.pdf, [retrieved: 09, 2017].

[6] www.blockchain.info, “Confirmed transactions per day”, [Online], Available: https://blockchain.info/charts/n-transactions [retrieved: 09, 2017].

[7] “Peercoin - Secure & Sustainable Cryptocoin.”, edited 2018, [Online], Available: https://peercoin.net/, [retrieved: 02, 2018].

[8] "PoET 1.0 Specification”, edited 2017, [Online], Available: https://sawtooth.hyperledger.org/docs/core/releases/latest/architecture/poet.html, [retrieved: 02, 2018].

[9] K. Olson, M. Bowman, J. Mitchell, S. Amundson, D. Middleton, and C. Montgomery, "Sawtooth: An Introduction", 2018, [Online], Available: https://www.hyperledger.org/wp-content/uploads/2018/01/Hyperledger_Sawtooth_WhitePaper.pdf, [retrieved: 02, 2018].

[10] D. Hopwood, S. Bowe, T. Hornby, and N. Wilcox, "Zcash Protocol Specifcation", edited 2018, [Online], Available: https://github.com/zcash/zips/blob/master/protocol/protocol.pdf, [retrieved: 02,2018].

[11] V. Buterin, “A next-generation smart contract and decentralized application platform”, 2013, [Online], Available: https://github.com/ethereum/wiki/wiki/White-Paper, [retrieved: 09, 2017].

[12] Iroha, “White Paper”, 2016, [Online], Available: https://github.com/hyperledger/iroha/blob/master/docs/iroha_whitepaper.md , [retrieved: 09, 2017].

[13] Ripple, “Solution Overview”, 2013,[Online], Available: https://ripple.com/files/ripple_solutions_guide.pdf, [retrieved: 09, 2017].

[14] “Welcome to Hyperledger Fabric”, edited 2017, [Online], Available: https://hyperledger-fabric.readthedocs.io/en/latest/, [retrieved: 09,2017].

[15] “The Monax Platform”, edited: 2017, [Online], Available: https://monax.io/platform/, [retrieved: 09, 2017].

[16] “Openchain - Blockchain Technology for the Enterprise”, edited: 2017, [Online], Available: https://www.openchain.org/, [retrieved: 09,2017].

[17] G. Greenspan, “MultiChain Private Blockchain — White Paper”, 2015, [Online], Available: https://www.multichain.com/white-paper/, [retrieved: 09, 2017].

[18] N. Szabo, “Smart Contracts: Building Blocks for Digital Markets,” 1996.

[19] G. Greenspan, “Why many smart contract use cases are simply impossible”, 2016, [Online], Available: https://www.coindesk.com/three-smart-contract-misconceptions/, [retrieved: 09, 2017].

[20] “Involver – Social Volunteering”, 2015, [Online], Available: www.getinvolver.com, [retrieved: 09, 2017].

[21] A. Marczewski, “Gamification: a simple introduction”, 2013, [Online], Available:

https://books.google.se/books?id=IOu9kPjlndYC, [retrieved: 09, 2017].

[22] K. Huotari and J. Hamari, "Defining gamification: a service marketing perspective", in Proceeding of the 16th International Academic MindTrek Conference, ser. MindTrek '12. New York, NY, USA: ACM, 2012, pp. 17–22, doi:10.1145/2393132.2393137.

[23] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining gamification", in Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, ser. MindTrek '11. New York, NY, USA: ACM, 2011, pp. 9–15, doi:10.1145/2181037.2181040.

[24] "Duolinguo", edited 2017, [Online], Available: https://www.duolingo.com/, [retrieved: 09, 2017].

[25] "Fitocracy", edited 2017, [Online], Available: https://www.fitocracy.com/, [retrieved: 09, 2017].

[26] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? – a literature review of empirical studies on gamification", in the 47th Hawaii International Conference on System Sciences, Jan 2014, pp. 3025–3034.

[27] M. H. Phillips and L. C. Phillips, "Volunteer motivation and reward preference: an empirical study of volunteerism in a large, not-for-profit organization", SAM Advanced Management Journal, 2010, pp. 12–19.

[28] Wikipedia, the free encyclopedia, "Wikipedia:Barnstars", 2004, [Online], Available: https://en.wikipedia.org/wiki/Wikipedia:Barnstars, [retrieved: 09, 2017].

[29] "Foursquare", edited 2017, [Online], Available: https://www.foursquare.com, [retrieved: 09, 2017].

[30] "X Box Live", edited: 2017, [Online], Available: https://www.xbox.com/en-US/live, [retrieved: 09, 2017].

[31] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Steering user behavior with badges", in Proceedings of the 22Nd International Conference on World Wide Web, ser. WWW '13. New York, NY, USA: ACM, 2013, pp. 95–106, doi:10.1145/2488388.2488398.

[32] M. Montola, T. Nummenmaa, A. Lucero, M. Boberg, and H. Korhonen, "Applying game achievement systems to enhance user experience in a photo sharing service," in Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era, ser. MindTrek '09. New York, NY, USA: ACM, 2009, pp. 94–97 doi :10.1145/1621841.1621859.

[33] J. Hamari and V. Eranti, "Framework for Designing and Evaluating Game Achievements," in Proceedings of DiGRA 2011 Conference: Think Design Play, 2011, [Online], Available: http://www.digra.org/wp-content/uploads/digital-library/11307.59151.pdf, [retrieved: 09, 2017].

[34] "The solidity programming language", 2015, [Online], Available: https://github.com/ethereum/wiki/wiki/The-Solidity-Programming-Language, [retrieved: 09, 2017].

Wearable Eye Tracking for Multisensor Physical Activity Recognition

Peter Hevesi*, Jamie A. Ward†, Orkhan Amiraslanov‡, Gerald Pirkl§ and Paul Lukowicz¶
*†‡¶German Research Center for Artificial Intelligence, Kaiserslautern, Germany
†University College London
¶University of Kaiserslautern
§Ostbayerische Technische Hochschule Amberg-Weiden
* peter.hevesi@dfki.de, † jamie@jamieward.net,
‡ orkhan.amiraslanov@dfki.de, § g.pirkl@oth-aw.de,
¶ paul.lukowicz@dfki.de

***Abstract*—This paper explores the use of wearable eye-tracking to detect physical activities and location information during assembly and construction tasks involving small groups of up to four people. Large physical activities, like carrying heavy items and walking, are analysed alongside more precise, hand-tool activities, like using a drill, or a screwdriver. In a first analysis, gaze-invariant features from the eye-tracker are classified (using Naive Bayes) alongside features obtained from wrist-worn accelerometers and microphones. An evaluation is presented using data from an 8-person dataset containing over 600 physical activity events, performed under real-world (noisy) conditions. Despite the challenges of working with complex, and sometimes unreliable, data, we show that event-based precision and recall of 0.66 and 0.81 respectively can be achieved by combining all three sensing modalities (using experiment-independent training, and temporal smoothing). In a further analysis, we apply state-of-the-art computer vision methods like object recognition, scene recognition, and face detection, to generate features from the eye-trackers' egocentric videos. Activity recognition trained on the output of an object recognition model (e.g., VGG16 trained on ImageNet) could predict Precise activities with an (overall average) f-measure of 0.45. Location of participants was similarly obtained using visual scene recognition, with average precision and recall of 0.58 and 0.56.**

Keywords–Wearable sensors; Machine learning; Activity recognition; Feature extraction; Computer vision.

I. INTRODUCTION

Self organizing teamwork – when a task is not assigned to individuals but to a group of people – is a common phenomena in the professional sector. Examples of this occur in many situations, for example on construction sites, in rescue teams, or in hospitals. The current work is a development of our earlier contribution on eyetracking for activity recognition [1], and forms part of a wider project that aims to study teamwork in industrial settings, and how the interactions and collaborations of individuals can characterize such teams.

One possible approach for understanding a group's processes and structures is to recognize the activity of each person and detect signs of interactions between the team members. This low level information can be fed into higher level recognition modules, e.g., for semantic analysis [2] or collaboration recognition. Obtaining accurate information about an individual's activities in a real world environment is a challenging task, one which most likely requires a variety of different wearable sensors and sensing modalities. These sensors have to be unobtrusive, accurate and scalable (both in price and availability). As mobile eyetrackers become cheaper and more widespread, they might be used for purposes other than simply gaze tracking. In this work, we aim to expand their usage towards detecting physical activities.

Eyetracking provides a useful insight into a person's attention. Attention in turn can provide useful indications of that person's activity. Previous work has explored the idea of using mobile eyetrackers for activity recognition [3]. To date the vast majority of this research has concentrated on activities directly related to visual attention and cognition (reading, watching TV, etc.) [4][5]. Our current investigation uses eye movement information, alongside camera images, to recognize physical activities, specifically activities related to an assembly and construction task. In doing so, we do not intend to develop highly specialized methods, optimized to perform well only in the selected scenario, but rather showcase the possibilities of the approach by using a collection of "off the shelf" methods.

In our earlier work, we described how wearable eye tracking devices can be used to provide useful informations about physical activities that are outside the mainstream eye-tracking application domain [1]. Here we add to this in two important ways. Firstly, we deepen our evaluation methodology by using a wider range of features, a range of different window durations, and add an additional, post-classification, temporal smoothing filter. Secondly, we apply state-of-the-art computer vision methods to the task of classifying activity and location using only egocentric camera images. This adds an extra layer of context to the activity classification, and points towards a promising sensor-fusion based approach for future work in this area.

In Section II, we provide an overview of recent articles and methods in our research field. In Section III, we describe the performed experiments, sensor setup and the datasets including the ground truth labels used during the evaluation. Section IV contains the overview and in-depth description of our methodology. The results for activity recognition using eye data are presented in Section V. In Section VI, we discuss possible use-cases of the egocentric videos recorded by the eyetracker devices. In Section VII, we highlight open questions for future work and conclude our findings.

II. Related work

As sensing technology continues to shrink in size and cost, an increasing number of researchers are turning their focus towards sensor-based activity recognition. One prominent research direction is using sensors embedded in the environment. A popular application of this approach is towards assisted living and smart homes. Methods for detecting activities of daily living are presented, for example, in [6], [7] and [8].

In workplace scenarios, like the one examined in the current paper, the sometimes dynamic and mobile nature of a task makes wearable sensing the best option for capturing it. Many studies deploy distributed body-worn or mobile inertial sensors to recognize a wide-range of physical activities (see [9] for an overview).

Sound is another common sensing modality. In [10], Lu et al. introduce a mobile-phone based system for classifying ambient sound, voices and music. Previous works use multiple streams of audio to recognize social situations [11][12], or to infer collocation and social network information [13].

Combined sound and movement data obtained from the smartphones of groups was used to analyses pedestrian congestion at busy thoroughfares, making use of changes in people's step-intervals and ambient audio [14]. Wrist-worn microphones and accelerometers were first used together to detect hand-tool activities in a wood workshop scenario [15]. More recently, these sensors were used to recognize physical collocation and collaboration of co-workers performing a group task [16].

A. Eye-based activity recognition

Eye tracking is a widely used technique in human computer interaction (HCI). It can be used, for example, in assistive technologies for people with limited motor skills (e.g., by Barea et al. [17]). Typically, researchers are interested in the object of a user's gaze – what it is that the user is looking at – e.g., in areas such as marketing research [18], or user interface design [19], [20], [21]. A different approach is to analyze the patterns created by eye movement in various situations. To detect reading activities while walking, Bulling et al. used patterns of eye fixation and saccadic movement recorded from changes in the eye's electrical activity (electrooculography, or EOG) [3]. Later, this work was extended to detect activities such as writing, reading, watching a video, etc. [22]. An advantage of a pattern-based approach is that no calibration is needed with a worldview video.

Changes to the blink rate can indicate different mental loads corresponding to different types of tasks. In [23], the authors describe methods for eyelid position and blink detection. Platforms like Google Glass include the ability to record blink rate, which when combined with head movement can be an effective method for recognizing activities [5].

Vidal et al. introduced a calibration-free, gaze interaction method based on tracking the smooth pursuit movements that occur when the eye follows a moving target [24]. And in [25] a commercial, wearable EOG system, the Jiins Meme, was used as a novel gestural input device based on a similar approach.

Shiga et al. proposed a system based on eyetracker and first person videos to recognize daily activities [26]. This work is the closest to our research. However, the authors focus on activities directly related to gaze (e.g., reading, video watching), whereas our work is more concerned with using eye-patterns to recognize physical activities that do not necessarily involve direct gaze (e.g., using a screwdriver).

B. Computer Vision on First Person Videos

Advances in computer vision, especially in the area of deep neural networks, have opened up a wide-range of possibilities. Frameworks like Keras [27], or Detectron [28], make it relatively easy for researchers to apply state-of-the-art vision algorithms quickly. Deep learning architectures optimized for recognizing images (e.g., [29] or [30]) are often available with pre-trained weights, and can be fine-tuning using only a few samples. Amos et al. presented a general purpose face detection framework [31] with an accuracy of ca. 93%. These methods usually support GPU-acceleration and can be used for real time video processing too.

Based on a day-long, first-person video footage of trips to an amusement park, Fathi et al. analyzed social structures of groups and the interactions within them [32]. Ryoo et al. use first-person perspective cameras to try and understanding how different people interact with an observer, and to be able to differentiate between friendly and hostile actions [33]. And in [34], the authors argue that to detect activities of daily living, the objects seen and interacted with could play a major role. Motivated by these results, we investigate the usefulness of egocentric videos to detect physical activities and to recognize locations in our dataset.

III. Experiment

To evaluate our different sensors and algorithms, we designed a construction-work-inspired, data-collection experiment, that was initially described in [16]. In the experiment, groups of up to four people work together on a demanding physical task (build a large TV wall), but crucially are free to choose how they go about this - and whether or not to collaborate with one another. Some of the sub-tasks can be performed in parallel, while others must be done in sequence. This led to a complex, highly variable, and noisy, dataset that closely-mirrors real-world construction scenarios.

A. Scenario

Four participants collaborate to build a 2.5 meter high TV wall consisting of 8 large LCD screens, 3 base panels, 18 screen spacers, and more than 50 screws. The parts are stored in containers at a storage area, which is separated by a ca. 25 meter long hallway from the assembly area.

The building phase included the following main steps: 1.) Unload screens (each screen weights 8 kg.) and other TV parts from the containers, 2.) Carry items to the assembly area, 3.) Assemble and place base items, 4.) Lift screens onto the wall, 5.) Fix screens on the wall by tightening the screws. After the build phase and a short break the participants perform the process in reverse: 6.) removing the screws, 7.) taking down the screens and other parts carefully, 8.) carrying back to the storage area, 9.) put them back into the containers.

Generally, the participants had the freedom to organize and execute the tasks as they thought it is best. Since the TV screens were quite heavy, they divided themselves almost every time into groups of two to carry and lift the screens. After the first components were delivered to the assembly area, they had the option to start with mounting and fixing the screws parallel to the transportation task. The overall task takes usually from 40 minutes up to 1 hour.

B. Wearable sensors

While performing the tasks, the participants were equipped with a with a mobile eyetracker, a sound recording device with two separate microphones and three inertial measurement units. This setup is shown in Figure 1.

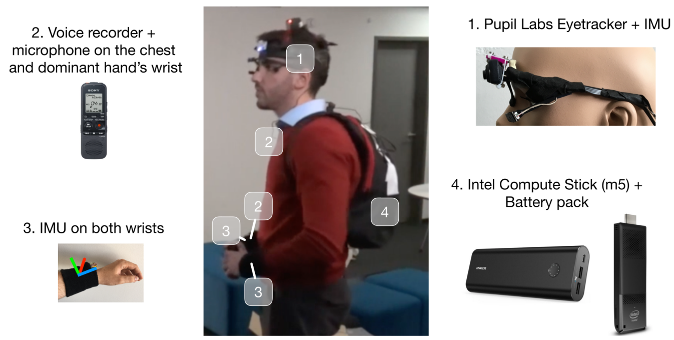


Figure 1. Recording setup for each participant includes an eyetracker connected to a small recording computer. Additional sensors: IMU on both arms and head, microphone on the wrist and at the chest.

a) Inertial measurement unit (IMU): For tracking movements of the participants, they wore IMUs on both wrists and one on the head. The IMU devices record 3-axis acceleration, gyroscope, and magnetic field as well as 3D orientation at approximately 40 Hz.

b) Sound recorder: Each participant wore two microphones: one on the dominant hand's wrist and a second one attached on the chest. The microphones were connected to a voice recorder capable of recording stereo sound and were saved as the two channels of an audio file.

c) Mobile eyetracker: In our experiment, we used Pupil Labs eyetracker devices (as described in [35]) with the 100 degrees field of view lenses for the world camera to cover more of the world's scene. The eyetracker setup additionally consists of an Intel Compute Stick with an m5 1.6 GHz processor (running Ubuntu 16.10) as a recording device for each person. They were powered by a portable 20100 mAh battery. This setup is able for mobile recording for ca. 1 hour 30 minutes, before the battery have to be recharged. The recording itself was done using Pupil Capture (v0.82) software. We implemented scripts to remotely control and monitor the recordings. The overall cost of this eyetracker setup is around 1600 Euros, which is significantly lower then many other commercially available mobile eyetracker solutions.

During our experiments, we observed a lot of issues with the eyetracker calibration. The main reason for these was the displacement of the eyetracker's frame on the head due to sudden movements and sweating as the participants performed heavy physical work. In an attempt to overcome these issues, we integrated the eyetracker's components into ski-goggles, which usually sit really tight and fixed on the face. However, this concept showed improved calibration stability, they were also more bulky and therefore not suitable for real world experiments. Also, for reproducibility reasons, we later decided to use the unmodified version.

C. Datasets and Labels

For our testing and evaluation, we used two recordings of the above described experiment performed by two different groups of participants. In each dataset, four stationary cameras recorded the scene additionally to the above mentioned wearable sensors (eyetracker, IMU data, sound recordings). Two of the cameras were recording the assembly area, one the storage area and one the hallway. The main purpose of these cameras is to help the annotation process. An important step was the synchronization of the signal sources and videos. This was done in a post processing step with the help of predefined synchronization gestures at the beginning of the experiment.

The data for each participant was annotated into 6 different activity events (adjust, screwdriver, drill, carry, screen placement, walk) and no activity (NA). These activities were then sub-divided for the 2-class (Large vs. Precise activities), and 1-class (Precise) analyses (details below). To evaluate how well the eyetracker can detect location and co-location of the subjects, we additionally labeled each participants positions throughout the datasets.

The degree of freedom to organize and perform the experiment resulted often in unexpected event flows with many short interruptions and activity changes. This proved to be a challenge to label, making low-level event annotation nearly impossible. On the other hand, this makes the data realistic. By keeping this in mind, we consider each ground truth label as a rough description of what a participant is mainly doing within a given time interval (from a few seconds up to a minute). Short interruptions (e.g., person taking additional screw from the desk or interacting with other participants) are not represented in this ground truth. In total, we labeled 606 activity events with an overall length of ca. 260 minutes. An additional 'No Activity (NA)' class was annotated to cover all the instances where a person is not doing any of the defined activities.

a) Six class problem: The detailed label set includes six classes (alongside the NA class):

1) Adjust: during these activities the subject is interacting (placing, taking or adjusting) with screws without any tool.
2) Screwdriver: subject tightens or loosens screws using screwdriver.
3) Drill: events when a participant tightens or loosens screws using a power drill with screwdriver attachment.
4) Carry: the times when one or two participants carry the heavy TV screens to or from the assembly area.
5) Screen placement: segments where screens are taken out of or placed back into the container or put on or taken off the TV wall.
6) Walk: person moves between assembly area and storage area (without carrying heavy objects).

b) Two class problem: With this label set, we want to investigate the performance of the system for separating actions involving large, location changing motions and subtle action requiring subtle movements. Thus, we defined two classes (in addition to NA):

1) Large: all events containing the above categories carry, screen placement and walk.
2) Precise: a combined set of the above defined adjust, screwdriver and drill categories.

c) One class problem: The third set looks only at the single class of Precise activities, as defined above, against a catch-all class comprised of everything else.

d) Location labels: With the current dataset, it would not be viable to annotate accurate 2D or 3D position of the participants. Instead, we created contextually meaningful location labels such as the person is at the storage area, on the hallway, on the corridor or at the assembly area. The layout of the experiment area including the location labels is showed in Figure 2. Based on these location labels, we derived the co-location of participants. Co-location is defined here as a pair of participants is in the same region (has the same location label).

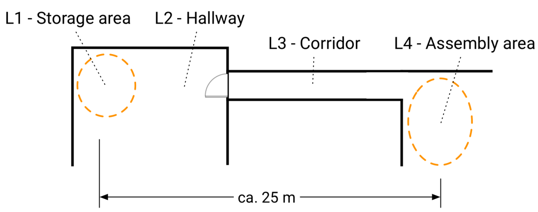


Figure 2. Floor plan of the experiment area with the highlighted location labels.

IV. Activity Recognition with Eye Tracking Data

This section describes our workflow and methodology used during the study. All code for processing the data is implemented in Python with Numpy.

A. Overview

The exploration space to find good parameter combinations is large. For example, we wanted to test different window sizes for the feature extraction on different label sets. We generate over 3000 feature samples per person per signal source for each dataset. To calculate one sample, the method selects raw values inside a window and performs the feature computations. Depending on the window size and the sampling rate of the data source, this window can contain more than thousand raw samples. So, recalculating a feature set can take longer time periods even on modern processors (usually between 5-20 minutes depending on the signal source). To be able to perform tests more efficiently and quicker, we divided our process into two main modules: 1) feature generation, 2) activity recognition. After a general description of these modules, we describe each important step in more detail.

1) Feature extraction module: As a first step, this module can find and parse the different data sources such as eyetracking data, measurements of the inertial measurement units (IMU) or the sound recordings. It can also load the files containing the ground truth information.

The feature extraction part takes the appropriate input signals and calculates the selected features (as described below) in the sliding windows. As a result, we obtain a table for each person, where each row contains the sample's time, the current activity label and the feature values.

An overview of the module's data flow is illustrated on Figure 3. We run this process for each parameter set (e.g., window size 5s, 6 classes, eye data only) and store the resulting feature matrices. Because of the high amount of data and number of combinations, this process usually has a longer runtime (for all combination it can take up to 7 hours). However, it needs only to be run once as long we do not introduce new features.

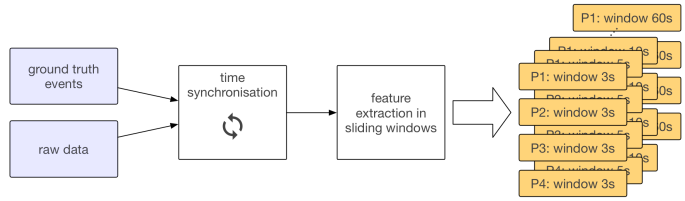


Figure 3. We first extract and synchronise the raw data, and then implement a parallelized process to evaluate different parameters (such as window size) for the selected features.

2) Activity recognition module: This module is designed to quickly evaluate feature sets and different parameter combinations. Figure 4 shows the flowchart of the module.

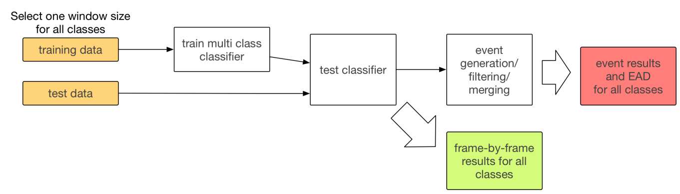


Figure 4. For each selected parameter set (window size, train-test split method, label set), we apply this evaluation process. As a result, we obtain frame-by-frame as well as event based performance metrics.

At first, it loads the previously generated feature tables for a selected window size. These are then split up into training and test sets based on the selected evaluation strategy. The training set is then used to train a classifier. After that, we test the classifier's frame-based performance by comparing the for the test set's features predicted labels with their ground truth labels.

Additionally, the frame by frame prediction results are then transformed into events. The event generation method can apply temporal smoothing for merging events close to each other and removing events that are too short. In the last step, we calculate event level metrics. The module can produce visualizations for both frame-based and event-based results.

B. Synchronization

During the recording, due to technical reasons, each device recorded the measurement values with timestamps based on their system time. The inertial measurement units and the eyetrackers use POSIX timestamps, however, there can be offsets between the current system times. Drift during the experiment is negligible, since all of the systems use quartz oscillators internally. In case of the stationary cameras and voice recorders, each frame's or sample's time is registered as the elapsed time from the beginning of the recording.

In the synchronization step, the goal is to bring the different signal sources to a common time base. For that purpose, as previously mentioned in Section III-C, each participant performed a synchronization gesture at the beginning of the experiment. The gesture was defined as jumping and clapping with the hands in front of the recording camera. Patterns of this gesture can be easily pinpointed on all of the raw signals. For the eyetrackers (first person video) as well as for the stationary cameras, we looked up the corresponding frames' timestamps.

Based on the known times of the gestures for each source, we could calculate the offset with the following simple equation:

$$\Delta t^{A-B}_{offset} = t^{A}_{sync} - t^{B}_{sync} \quad (1)$$

where A and B are references to signal sources (e.g., P4's eyetracker and main video). t^{A}_{sync} and t^{B}_{sync} refer to the same synchronization event's time on respectively A or B's system time. With the known offset, we can convert times from one signal's time to the others using the following formula:

$$t^{A}_{i} = t^{B}_{i} + \Delta t^{A-B}_{offset} \quad (2)$$

By applying this method, we converted each source channel's timebase to the main video's time for easier referencing.

C. Feature Extraction

All of the below described features are calculated with the sliding window method over the evaluation time. This means that, we take the raw signal(s) between the start and end of the current window and calculate the source specific features. The window is then shifted forward until it reaches the end of the experiment. In this analysis, we use the center of the window to look up the activity label of the participants and also to combine feature rows from different sources (IMU, eye data or sound).

1) Eye features (eye): In each window, we calculated 14 statistical features on different eye related events and properties. These are:

- mean (μ) and standard deviation (σ) of the fixation length during the window
- μ and σ of the gap's duration between fixations
- μ, σ and zero-crossings (ZC) of the eye's spheric coordinates (θ and ϕ)
- μ and σ of the pupil size
- μ and ZC of eye position estimator's confidence value.

ZC, a simple measure of dominant signal frequency, is calculated by counting the zero-crossings on each window after subtracting μ.

Based on our initial findings, we expect the length of the fixations and the gap between two fixations to be a good indicator for increased attention. Information about the 3D orientation of the eyeball can help us to distinguish between different type of eye activities. The pupil size could help to distinguish dark and bright environments. Accommodation (change of viewing distance) can also cause changes of pupil size. Changes in the confidence of the eye position estimator correlate usually (if no displacement, loss of calibration or other problems) with the blink rate of the person. Blink rate can vary very much from person to person, but can also be related to specific type of activities (e.g., when focusing, blink rate drops).

We calculated eye features using window sizes of 3, 5, 10, 15, 30, 45 and 60 seconds.

2) Accelerometer features (ACC): Only the accelerometer signals (ACC) from each person's right-wrist IMU are used in this study. These 3-axis accelerometer signals (x,y,z) are combined to give a single orientation-invariant reading by the formula:

$$a = \sqrt{x^2 + y^2 + z^2} \quad (3)$$

For each of these readings five standard features are calculated across a 1 second rolling window, these are: mean (μ), standard deviation (σ), short-term energy (E), zero-crossing rate (ZC) and skewness (γ).

3) Sound features (snd): Sound signals from each participant's dominant wrist, s_r (all were right-handed), and head, s_h, are downsampled from the recording rate of $44.1kHz$ to $8kHz$ (16 bit). In the first step, two features are extracted for each of these across a rolling window of 40 milliseconds: short-term energy, E, and zero-crossing rate, ZC. These features were chosen because of their widespread use in low-cost speech and audio analysis [36]. In a second step, we re-sampled these features using a window of 1 second for smoothing.

4) Fusion of features (acc+snd and all): One of our goals is to evaluate how well a combination of different data sources performs on the task of detecting physical activities compared to the baseline performance. For this purpose, we use the approach of combining feature matrices and training a new classifier with the merged matrix.

When merging feature matrices, we look up the corresponding row of the new matrix for each row of the original matrix by selecting the one with the closest sample time. If no feature row can be found with a smaller sample time difference than a threshold of one second, we skip this row (new feature cells have then non-valid values).

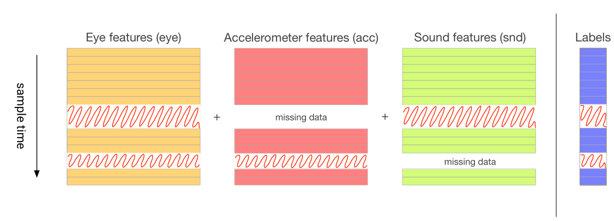


Figure 5. Merging of feature matrices. To eliminate the issue of missing data blocks in different sources, we simply remove all rows from the feature matrix and label vector, which consist non-valid values.

One of the challenges here is that in many cases data is not available or corrupt for a time period on one of the sources. Figure 5 shows the selected approach to handle missing data. When any row of the merged feature matrix includes any corrupt or non-existing values, we simply remove this line from the matrix. This also means that these rows are not considered in the evaluation.

In this work, we used combinations of:

1) ACC and sound features for comparing eye features with other modalities (acc+snd)
2) eye, ACC and sound features to see, if the eyetracker data can improve the overall recognition rate (all)

D. Classifier training

For all evaluations, we used a Naive Bayes classifier with a One vs. Rest training strategy. This means training one classifier for each class, where positive samples are taken from training instances of the current class, and negative samples from all other classes. This strategy typically achieved 1-2 percent better results than the One vs. One strategy that we used in our earlier work [1]. We also experimented with different classifier methods, but found the Naive Bayes to be sufficient for the purposes of the current work. All methods for training and testing are implemented on Python using the scikit-learn toolkit [37].

E. Event generation

In the prediction phase, the classifier produces one label per sample (row of the feature matrix). We refer to these as frame-based or frame-by-frame predictions. When the same label is returned in a sequence, we can combine them into an event by saving the time of the first occurrence as the event's start and the last as the event's end. If there is only one single frame in the sequence detecting a class, we use the sample distance as the event's length to avoid zero length events, which is 1 second in our case.

We implement a temporal smoothing on event level to reduce the number of false detections. In a first step, if an event follows a previous event of the same class within a specified time, they will be merged into one event. We used a threshold of 10 seconds for the event merging. Finally, events with a duration smaller than a threshold of 1.5 seconds are removed. Figure 6 shows the event generation and filtering concept.

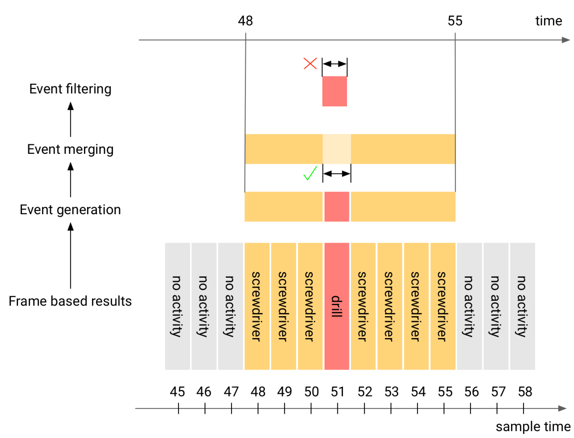


Figure 6. Event generation process. Frame-based results are transformed into events, events following each other in a close distance can be merged and very short events can be removed.

F. Evaluation

1) Train-test split: By using dataset A and B, we defined four evaluation runs, two for dataset dependent and two for dataset independent. Dependent 1 strategy uses only data from experiment A for training and testing. Dependent 2 does the same on dataset B. For independent 1 strategy, we use dataset A for training and dataset B for testing. In case of independent 2, the classifier is trained on data from B and tested on data from A.

a) Experiment dependent evaluation: For experiment dependent evaluation, data from one dataset (dataset A or B) were split into training and test sets by the leave one person out method. For example, data of P1, P2 and P3 participants were used to train the classifier and data of P4 to test it. The method is therefore person independent. The method was applied for all four combinations of an experiment independently and the results were averaged. With this approach, the test data is always unseen for the classifier and at the same time we get an estimation about the generalization error.

b) Experiment independent evaluation: A second approach for the evaluation is to use one dataset for training and the other one for testing. Both for training and test, we include data from all participants of the corresponding dataset. These tests usually indicate how well the system can generalize the results and handle later datasets without any additional training effort.

2) Performance Metrics:

a) Frame based evaluation: Each item of the classifier's prediction output is compared to the corresponding ground truth labels for the frame based evaluation. For that, we use the standard convention, where each predicted label is considered as true positive (TP) if its equal to the samples ground truth label or as false positive (FP) otherwise. A ground truth label is a false negative (FN) if the predicted label for the same sample is different.

We calculate then precision and recall values as defined by the standard definition. That is for precision:

$$P = \frac{TP}{TP + FP} \tag{4}$$

and for recall:

$$R = \frac{TP}{TP + FN} \tag{5}$$

Accuracy score is calculated by:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

with TN for the true negatives.

b) Event based evaluation: In the event based evaluation, we compare detection events with the ground truth. A detected event is considered as a true positive (TP_{det}) if it has an overlap with a ground truth event of the same activity (for the same participant) or as a false positive (FP_{det}) otherwise. Similarly, ground truth events are labeled as true positives (TP_{gt}) if they are detected at least once otherwise as false negatives (FN_{gt}).

We calculate the event-based precision (P) and recall (R) analogous to the standard frame-based definitions, with the addition of the events' durations as weight factor:

$$P_e = \frac{TP_{det} \cdot \Delta T_{det}^{TP}}{TP_{det} \cdot \Delta T_{det}^{TP} + FP_{det} \cdot \Delta T_{det}^{FP}} \tag{7}$$

and,

$$R_e = \frac{TP_{gt} \cdot \Delta T_{gt}^{TP}}{TP_{gt} \cdot \Delta T_{gt}^{TP} + FN_{gt} \cdot \Delta T_{gt}^{FN}} \tag{8}$$

Additionally, we calculate scores specially designed for event based evaluation [38]. The approach is designed to acknowledge the fact that event-based recognition is not as clear-cut as true/false, positive/negative-based evaluations imply. Real-world recognition can be fragmented in time, or multiple events can be merged together. The evaluation methodology used works by labeling both ground truth events and prediction events into separate error categories.

Error categories that are applied to the ground truth include:

- Deletions (D): when a ground truth event is not detected at all, practically the same as a false negative
- Fragmented (F): a ground truth event is detected by multiple (at least 2) true positive fragments
- Fragmented and Merged (FM): the ground truth event is detected in multiple (at least two) fragments, and at least one of the detections also overlaps with another ground truth event
- Merged (M): a detection event covers more than one (true positive) ground truth events

Error categories that are applied to the detected events include:

- Insertions (I'): equivalent of false positives, when a detection event has no overlap with a ground truth event at all
- Merging (M'): detection events that overlap with more than one ground truth events
- Fragmenting and Merging (FM'): a detection event, which overlaps with at least two ground truth events, where one of them at least is also covered by another detection event
- Fragmenting (F'): when an event is detected during a ground truth event that is also covered by other detection events

Finally, one category can be applied to both ground truth and detected event output:

- Correct (C): a ground truth event corresponds to one true positive detection event

In our tests, we used the open source Python implementation of this scoring mechanism and associated metrics [39].

V. RESULTS

This section contains the summary of our most important results regarding the usage of eye movement features and it's combination with acceleration and sound features for physical activity recognition.

A. Window size for eye features

In a first step, we evaluated the influence of the eye features' window size on the classification results. Features for each person in both datasets were generated using eight different window sizes ranging from 3 seconds to 60 seconds. We performed tests with these features on all three label sets (one-, two-, six-class problem) and using all four evaluation strategies.

Results for the window size evaluations for the different label sets are presented on Figure 7. The y axis on these figures represents the accuracy value as defined in Equation (6). The six class problem's results are displayed on Figure 7-A. In this case, the overall accuracy seems to be very similar across different evaluation strategies (dependent, independent), especially at small window sizes. With an increasing window size, two of the strategies perform worse. For the poor performance

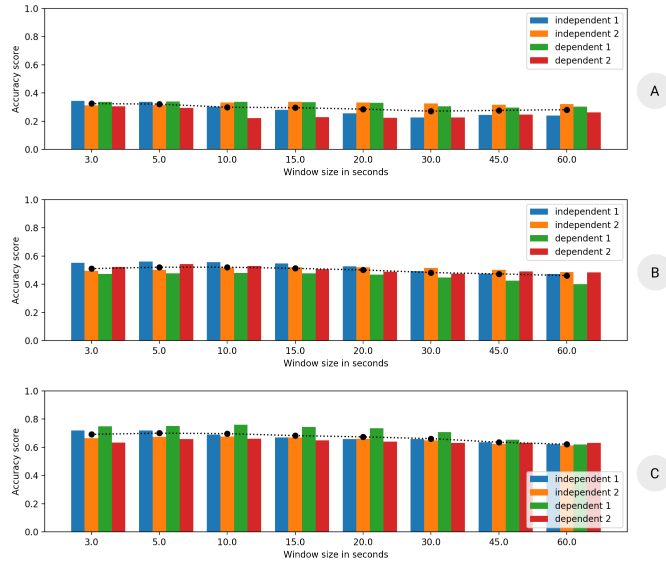


Figure 7. Activity recognition accuracy scores (frame-based) for eye features calculated with different window sizes on the six class problem (A), on the two class problem (B) and on the one class problem (C).

in independent 1 and dependent 2, probably the missing data points in dataset B are responsible.

For the two class problem (Figure 7-B) and the one class problem (Figure 7-C), the different strategies perform very similarly to each other. As going from the six class problem towards the one class problem, as expected, the overall accuracy of the classifier improves significantly and is always better than random guessing. The average accuracy across strategies for the ten second window is for the six class problem 0.37, 0.56 for the two class and 0.71 for the one class problem.

In the further analysis, we use the ten second window size to extract eye features, because it performed fairly good on each label set. Larger window sizes tend to be more sensitive about the distribution of the training and test data as we could observe on Figure 7-A.

B. Results for six classes

The confusion matrix on Figure 8-A shows the frame based results using the eye features extracted in ten seconds windows. The classifier was evaluated using the independent 1 strategy. However, single classes not recognized very well (but still better than random guessing), they point out two major groups in the data. This two groups are defined as Large and Precise class in the two class problem.

Figure 8 shows the results using the same strategy, but with the combination of all features into a single feature matrix. The results are still not perfect, but the improvement is visible and the two categories can still be observed.

The results on Figure 7 show us that different evaluation strategies perform similarly. However, it varies from label set to label set, which one is better, the other results are typically in a range of ± 5 percent. That is why and for a better overview, we decided to provide in the following the performance values only for the independent 2 strategy.

Table I shows all frame-based results of the six class problem. For each feature set (eye: using only eye based

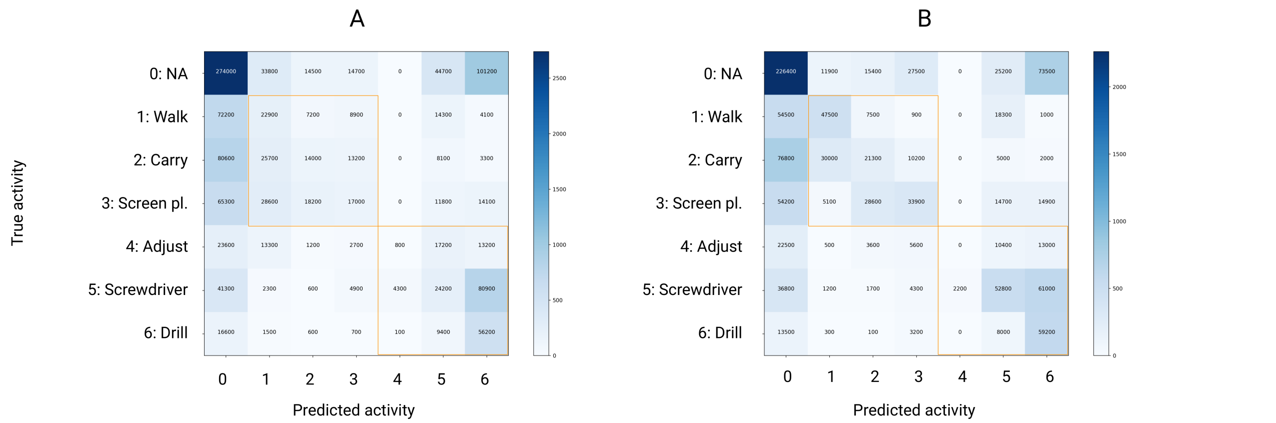


Figure 8. Confusion matrices for six the class problem using eye-based features only (A) and using all features (B) with 10 seconds window for dataset independent training 1. These frame-based results implicate that some classes should be merged together (see two class problem).

TABLE I. FRAME BASED PRECISION AND RECALL VALUES FOR THE SIX CLASS PROBLEM, DATASET INDEPENDENT TRAINING

class	*eye*		*ACC+snd*		*all*	
	P	R	P	R	P	R
no activity	0.45	**0.63**	0.44	0.44	**0.47**	0.60
Walk	0.32	0.04	0.48	**0.60**	**0.49**	0.42
Carry	**0.32**	0.15	0.23	0.04	0.27	**0.19**
Screen pl.	0.14	0.09	0.27	0.01	**0.40**	**0.29**
Adjust	0.00	0.00	0.00	0.00	0.00	0.00
Screwdriver	0.26	0.29	0.35	0.23	**0.39**	**0.33**
Drill	**0.32**	0.14	0.12	0.62	0.26	**0.70**

features, ACC+snd: combination of accelerometer and sound features, all: combination of all three feature sources), it presents precision (P) and recall values (R).

One important observation is that for almost every class, the combined feature set (all) outperforms the other two baselines. In some cases (Screwdriver, Screen placement), the combination is clearly better for both recall and precision values. In contrast, the Walk activity's recall is higher (0.60) using ACC+snd features and is reduced in combination (0.42). Recognition of the Adjust activity does not work, most probably, because it is not well represented in the training samples (due to missing data, the activity's sample instances were cut out by chance). In some other training-test split strategies, it can be detected (as seen on Figure 8-A).

TABLE II. EVENT BASED PRECISION AND RECALL VALUES FOR THE SIX CLASS PROBLEM, DATASET INDEPENDENT TRAINING

class	*eye*		*ACC+snd*		*all*		*filter*	
	P_e	R_e	P_e	R_e	P_e	R_e	P_e	R_e
Walk	0.54	0.45	0.46	**0.92**	0.53	0.67	**0.67**	0.54
Carry	0.52	0.48	0.19	0.40	0.39	**0.66**	**0.68**	0.51
Screen pl.	0.42	0.52	0.22	0.09	0.50	**0.61**	**0.51**	0.41
Adjust	0.10	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Screwdriver	0.19	0.53	0.40	0.88	0.42	**0.90**	**0.49**	0.78
Drill	0.22	0.61	0.14	**1.00**	0.26	0.61	**0.30**	0.53

Results of the event-based analysis for the six class problem are presented in Table II. For the evaluation of the first three columns (eye, ACC+snd, all) no event filter or merging is applied. The fourth column named "filter" merges events that are closer than 10 seconds and removes events shorter than 1.5 seconds based on the combined (all) features' event detections.

Most important observations in the event-based results are:

- for some classes (carry, screen placement) eye features work best, while for other events (e.g., walk, screwdriver) the accelerometer and sound combination work best
- the merged features (all) achieves overall better results
- as expected, by using the event merging and filtering, the precision of the event classifiers could be increased (at the cost of a lower recall)

To understand the nature of the errors, we analyzed the results of the combined feature set and event filtering (filter) with the event analysis diagrams proposed by [38]. Figure 9 illustrates the detailed event score distribution for different classes for the dataset independent training. According to this analysis, 60% of Walk events are successfully detected. However, detected events for Screwdriver and Drill classes are mainly dominated by insertions (I'). These activities are easily mistaken for one another (as can also be seen in Figure 8), which can help account for the low precision values obtained, too.

C. *Results for two classes*

We expect the classifier to perform significantly better for the two class problem, as many of the errors reported above are due to confusions between similar activities (again, see Figures 8 A and B).

TABLE III. FRAME BASED PRECISION AND RECALL VALUES FOR THE TWO CLASS PROBLEM, DATASET INDEPENDENT TRAINING

class	*eye*		*ACC+snd*		*all*	
	P	R	P	R	P	R
no activity	**0.50**	0.43	0.43	**0.57**	0.48	0.52
Large	0.62	0.45	0.61	0.36	**0.69**	**0.48**
Precise	0.47	**0.74**	0.38	0.45	**0.55**	0.73

Frame by frame results of the two class problem are in Table III. The values in this table confirm our expectations. Using only data from the eyetracker (10 second window, independent training) Large class can be detected with a precision and recall of 0.62 and 0.45, meanwhile the classifier achieves 0.47 precision and 0.74 recall for Precise class. Similarly to the

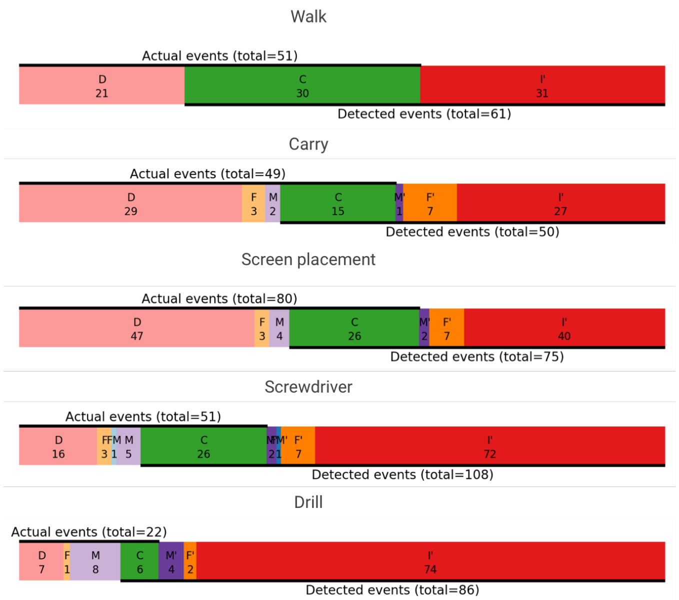


Figure 9. Event analysis diagram for the six class problem (class Adjust had zero detections), dataset independent training. (C denotes correct events, D deletions, I' insertions).

TABLE IV. EVENT BASED PRECISION AND RECALL VALUES FOR THE TWO CLASS PROBLEM, DATASET INDEPENDENT TRAINING

class	*eye*		*acc+snd*		*all*		*filter*	
	P_e	R_e	P_e	R_e	P_e	R_e	P_e	R_e
Large	0.74	0.83	0.54	**0.86**	0.80	0.72	**0.80**	0.61
Precise	0.57	0.96	0.39	**0.99**	0.62	0.98	**0.63**	0.76

six class problem, the combined feature set has an improved overall performance compared to the baselines.

Figure 10 presents a perfect example while the pure frame-based evaluation is not sufficient in our application. The temporal distribution of the detections is also very important. The figure shows the ground truth and detected events for person 2 over the dataset A. It can be observed that the majority of the detections for a class is around the real activity. Using event filters as described above, we also can filter out some of the wrong detections.

Table IV shows the actual event length weighted precision and recall values for all four categories (eye features only, using ACC+snd, using all features, using event filter as well). For Precise events, the classifier has a precision of 0.62 with a recall of 0.98 when using the combined feature set.

The event analysis diagram (EAD) of the two class problem is shown in Figure 11. For both Large and Precise classes many of the correctly detected events are merged with other ground truth events. This merging together of correct events might not be strictly incorrect for some applications, but it does indicate ill-defined event borders. The Precise class events perform slightly better in terms of number of correctly discovered events. However, this result is dominated by insertion errors, which are likely responsible for the lower precision value.

D. Results for one class

Instance of the Precise class are important for the high level analysis, since this activity category is typically performed alone. It is characterized by, small precise movements over a longer time period. In this evaluation, we test the classifier to spot these events against the empty class ('no activity').

TABLE V. FRAME BASED PRECISION AND RECALL VALUES FOR THE ONE CLASS PROBLEM, DATASET INDEPENDENT TRAINING

class	*eye*		*ACC+snd*		*all*	
	P	R	P	R	P	R
no activity	0.89	0.64	0.75	**0.74**	**0.90**	0.70
Precise	0.43	0.78	0.35	0.23	**0.50**	**0.80**

Results of the frame-based analyses are in Table V. As expected, using only eye based features results in better performance then using the accelerometer and sound combination. The best performance is produced again by the all combined feature set with a precision of 0.5 and recall of 0.8 percent.

TABLE VI. EVENT BASED PRECISION AND RECALL VALUES FOR THE ONE CLASS PROBLEM, DATASET INDEPENDENT TRAINING

class	*eye*		*ACC+snd*		*all*		*filter*	
	P_e	R_e	P_e	R_e	P_e	R_e	P_e	R_e
Precise	0.59	0.96	0.35	0.97	0.64	**0.98**	**0.66**	0.81

The event based results in Table VI show the same improvement characteristics as observed for the other label sets. However, the accelerometer and sound based recognition has a high recall, but a poor precision value for a single class problem. Using only the eye features results in a fairly good 0.59 and 0.96 precision and recall. These values are further improved when using the combined feature set with 0.64 precision and 0.98 recall. When applying event filters, we get higher precision (0.66) on cost of the recall value (0.81).

The event analysis diagram (EAD) (see Figure 12) of the one class problem shows less deletions and insertions for the Precise class compared to the two class problem's figure (see Figure 11). The majority of the ground truth events are detected merged with other events. The reason for this is probably, as seen on Figure 10, that many occurrences of the Precise class (ground truth) occur close to each other.

VI. COMPUTER VISION ON EGOCENTRIC VIDEOS

In addition to the eye-feature based activity recognition, we also explored the possibility of using the world camera images as an information source. Specifically, we wanted to find out how "off the shelf" machine vision tools might be used to automatically process the camera data. In this section, we describe our findings on the use of automatic face detection, object recognition around the gaze point, and scene recognition.

A. Objects of interest

The initial idea is based on the assumption that a person looks at specific objects more often if these are related to his or her current task. One such example is the activity of fixing screws with a screwdriver. We would expect that it is more likely to find a screwdriver in the egocentric video of the person doing this task, than it would be in the video of someone who was doing something else. If we can detect the objects that the person looked at during specific activities, we can train a classifier using this information and use it to later recognize the same activity again.

Figure 10. Ground truth events and detections for the two class problem for person two in dataset A without (top) and with (bottom) event filtering and merging. The classifier was trained using all features combined and only with data from dataset B.

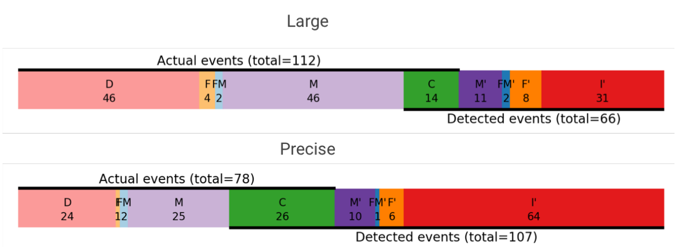


Figure 11. Event analysis diagram for the two class problem, dataset independent training

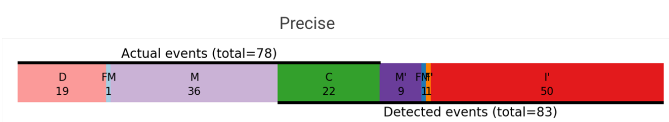


Figure 12. Event analysis diagram for the one class problem, dataset independent training.

To test the feasibility of this idea, we implement the following workflow:

1) Crop each frame of the eyetracker's world video around the gaze point with a radius of ± 200 pixels horizontally and vertically to obtain the region of interest (ROI).
2) Use deep learning models pretrained on 1000 classes of the ImageNet database to obtain prediction weights on the ROI. We tested the Keras implementation of the following models: InceptionV3, ResNet50, VGG16, VGG19 and Xception.
3) We calculated the average of the weights in one second windows over the dataset and stored as a feature matrix.
4) Using the activity ground truth label and the feature matrix generated from the prediction weights, we trained a Naive Bayes classifier with the same methods and training strategies as for the eye based features.

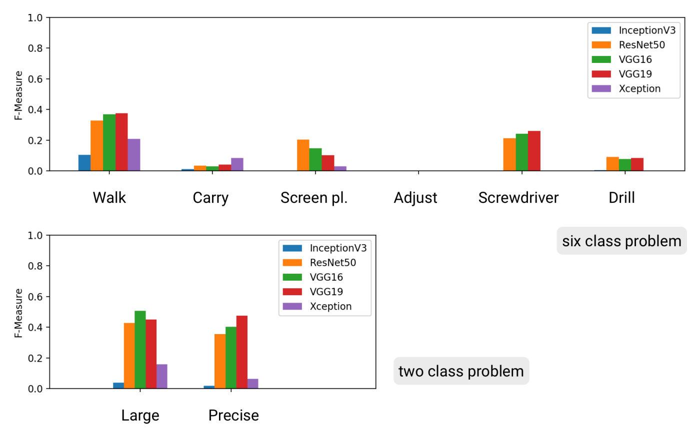


Figure 13. F-measure of activity classifiers trained on prediction weight features of different models with dataset independent 2 training for six class problem (top) and two class problem (bottom).

Results for the activity recognition using the prediction weight based features are presented on Figure 13 for the six class problem (top) and for the two class problem (bottom). The figure displays the f-measure, that can be interpreted as a weighted harmonic mean of the precision and recall, for each class. The values represented on the figures are from a dataset independent training. The dataset identical trainings showed nearly identical results without much improvement.

On Figure 13, we can see that using the ResNet50, the VGG16 or the VGG19 models, the classification result for Walk and Screwdriver are comparable with eye features or the all feature combination from above. InceptionV3 and Xception models are not working well on these datasets. Detection results for the two class problem (Figure 13) are worse than the performance of the eye features on the same two classes, but still better then random guessing of the class though.

Figure 14. Top ten classes for scene recognition on first person camera images at different locations using a VGG16 model trained on the Places365 dataset. As expected, the predictions are very similar, but the weights' distributions are slightly different.

Motivated by these results, we merged the object recognition features with the eye, accelerometer and sound features from earlier in the paper using the same merging method as before (Section IV-C). The merged features are then evaluated using the same strategies and concepts as described in Section IV-F.

Table VII and VIII contain frame- and event-based evaluation values for two feature combinations:

1) eye features extracted with a 10 second window together with object recognition features (eye+obj)
2) eye, ACC and sound features combined with the object recognition features (all+obj)

In both cases, we use the VGG19 pretrained model to obtain the object recognition features and use independent evaluation strategy 2. By comparing Table VII with the above results in Tables I and II, the values are slightly worse than before. For the two class problem, however, this approach brings a minor improvement on frame-by-frame as well as on event level for both feature sets (see Table VIII vs. Tables III and IV).

TABLE VII. EVALUATION RESULTS FOR COMBINATIONS WITH THE OBJECT RECOGNITION FEATURES USING THE PRETRAINED VGG19 FOR THE SIX CLASS PROBLEM, DATASET INDEPENDENT TRAINING

	Frame-based				*Event based*			
	eye+obj		*all+obj*		*eye+obj*		*all+obj*	
class	P	R	P	R	P	R		
no activity	**0.53**	0.33	0.52	**0.45**	-	-	-	-
Walk	0.25	**0.74**	**0.43**	0.44	0.48	**0.86**	**0.55**	0.77
Carry	0.25	0.02	**0.34**	**0.15**	0.37	0.07	**0.59**	**0.61**
Screen Pl.	0.21	0.2	**0.34**	**0.33**	0.33	**0.59**	**0.51**	0.56
Adjust 0.00	0.00	0.06	0.01	0.00	0.00	0.00	0.21	0.02
Screwdriver	0.24	**0.33**	**0.32**	0.27	0.42	0.77	**0.46**	**0.82**
Drill	0.08	0.12	**0.21**	**0.76**	0.14	0.47	**0.21**	**0.56**

TABLE VIII. EVALUATION RESULTS FOR COMBINATIONS WITH THE OBJECT RECOGNITION FEATURES USING THE PRETRAINED VGG19 FOR THE TWO CLASS PROBLEM, DATASET INDEPENDENT TRAINING

	Frame-based				*Event based*			
	eye+obj		*all+obj*		*eye+obj*		*all+obj*	
class	P	R	P	R	P	R		
no activity	**0.53**	0.36	0.52	**0.39**	-	-	-	-
Large	0.66	**0.67**	**0.69**	0.55	0.66	0.70	**0.83**	**0.80**
Precise	**0.53**	0.73	0.52	**0.84**	**0.63**	**0.97**	0.60	0.74

B. Face detection

For analyzing interactions in the group, an important information source could be if we are able to detect if the person looks on another participant's face. Additionally, co-location information can be important to recognize collaboration events. If we can detect specific person's face on the egocentric video, this can be a clear sign of co-location. As an initial exploration towards these goals, we tested the face detection library OpenFace [31]. The authors report an accuracy of ca. 0.93 with their default model on the LFW benchmark [40].

We run the face detection on each frame of the eyetracker videos an saved position, bounding box size and face features together with the corresponding timestamp into a feature file for later processing. Figure 15 illustrates some examples of the detected faces in a sequence.

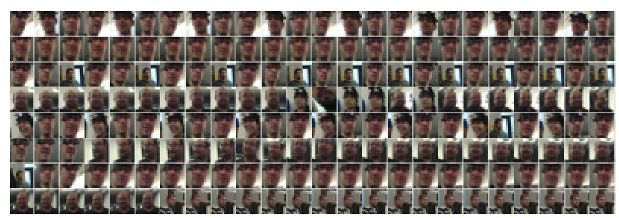


Figure 15. Examples for face detection using the OpenFace library on our dataset. It spots faces robustly throughout the experiment. (Image quality artificially reduced here).

We performed a statistical analysis of the temporal distribution of the face detections compared with activity labels. We found that during activities of the Precise category (using screwdriver, drill or adjust) the average number of face detections per second is much lower (0.05) than during Large activities (0.26). The reason behind this statistical difference is, that these activities are typically performed independently from other participants, meanwhile the persons are physically connected for many Large activities especially during the Carry event. Also, they used the time, e.g., during carry event for social interactions. Another interesting observation we made, was that two of the participants looked at each other much more often than other ones. We found that meanwhile these two knew each other well from before the experiment, they had no or little prior relations to the other ones, suggesting that this kind of data could be used for analyzing social structures within the group, as also done by [32]. But further analysis of this topic is outside the scope of the current work.

Lastly, we compared temporal distribution of the face detections against co-location of the participants. Figure 16 shows a scatter plot with the number of faces detected in a 15 seconds window against the ratio if the person is co-located with one or more other participants during this window. On the x-axis, zero means that during the window the person is completely alone, 100 percent means that during the whole time he or she is co-located with exactly one other participant.

The distribution of the sample points suggests a correlation between a high number of faces detected and co-location. However, this feature alone is not sufficient to predict co-location. Even if all four participants are at the same location, they do not necessarily look in each other's direction. This explains the sample points on the figure that have zero or very few faces detected even during higher values of co-location.

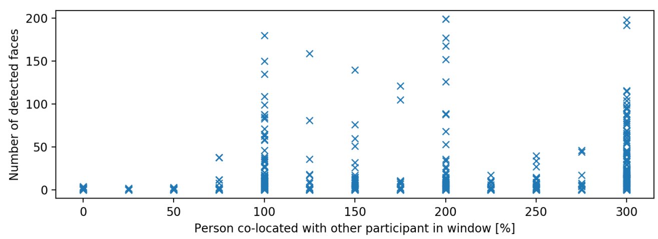


Figure 16. Number of faces detected in a window of 15 seconds versus the time spent co-located with another participants during this window. Values above 100% mean that the subject is co-located with more than one participant at the same time.

C. Location detection with scene recognition

Finally, we investigate location detection using egocentric video. Our assumption is that deep learning models trained to recognize different scenes can extract useful features for distinguishing locations in our dataset.

To test this assumption, we used the Keras implementation [41] of an VGG16 model trained on the Places dataset [42]. This model tries to predict a scene description for the input image. Figure 14 shows example predictions (top 10 for each) for three different locations in our dataset. As expected, the predictions are similar (are parts of the same building), but the order (and also the weight distribution) is different for each.

We calculated the output weights (predictions) of the network for each video frame from the eyetrackers and saved them into feature matrices. We also experimented to cut the network at an earlier stage (not the final class predictions), but that did not improved our results and introduced higher computational needs. On the generated feature set and location labels, we evaluated the concept with a nearest neighbor classifier using only one person's data for training and every other participant's data for testing. Precision and recall values were computed and then averaged for all four combinations of dataset A. For the four location classes, the classifier achieved and averaged precision and recall of 0.5774 and 0.558, well above random values. Considering the difficulty of the task (image quality, similarity of the places, in many cases bad lighting as visible on Figure 14), it is a promising result and could be an useful input for a high level decision making system.

VII. Conclusion and Future work

We evaluated the use of data from wearable eye trackers as part of a multi-sensor system for recognizing real-world physical activities. In the two datasets we used, participants perform (sometimes heavy) physical construction tasks in an unrestricted order over a duration of around 90 minutes. Due to the realistic data collection setup, data is in many cases unavailable or unreliable.

Despite the challenging conditions, we show in this work that it is still possible to obtain useful recognition results using eye tracking. Although the gaze point may not be correct, due to factors such as loss of calibration during the experiment, we can nonetheless successfully recognize key activity categories by utilizing calibration free eye features (based on non-gaze eye properties and movements). The results additionally show that recognition can be achieved in both a person and dataset independent way. That means that we can expect the system to work on future datasets without any additional training effort.

The combination of eye features with simple acceleration (measured on the right wrist), and sound features, produces even more reliable results especially on an event level. Additional event filtering can further reduce the number of false positive outputs. For detecting the broader classes of Large activities (involving full body movement) and Precise activities (involving smaller hand-tool manipulations), we achieved event-based (duration weighted) precision and recall values of 0.74 and 0.83 using only eye features, and 0.57 and 0.96 using the combination of eye, acceleration and sound features.

Exploration of the image processing methods on egocentric videos showed promising results despite the challenging dataset (bad lighting, blurry images caused by sudden movements, etc.). Although the initial results for location, co-location detection, or activity recognition using recognized objects are not perfect, they point towards potentially valuable areas upon which future research on this topic might be focused.

We aim to further explore the combination possibilities of eye features with visual object detection and scene analysis. Additional topics for investigation are the combination of sound analysis with face detection to recognize interactions within the group, and hand detection on egocentric videos for object manipulation detection.

Acknowledgment

The work is funded by the German Federal Ministry of Education and Research (BMBF) through the project iGroups.

References

[1] P. Hevesi, J. A. Ward, O. Amiraslanov, G. Pirkl, and P. Lukowicz, "Analysis of the usefulness of mobile eyetracker for the recognition of physical activities," in *UBICOMM 2017 The Eleventh International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. IARIA XPS Press, November 2017, pp. 5–10.

[2] A. Gruenerbl, G. Bahle, and P. Lukowicz, "Detecting spontaneous collaboration in dynamic group activities from noisy individual activity data," in *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE, 2017, pp. 279–284.

[3] A. Bulling, J. A. Ward, H.-W. Gellersen, and G. Tröster, "Robust recognition of reading activity in transit using wearable electrooculography," in *PERVASIVE*, May 2008, pp. 19–37.

[4] M. Vidal, D. H. Nguyen, and K. Lyons, "Looking at or through?: Using eye tracking to infer attention location for wearable transparent displays," in *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, ser. ISWC '14. New York, NY, USA: ACM, 2014, pp. 87–90. [Online]. Available: http://doi.acm.org/10.1145/2634317.2634344

[5] S. Ishimaru, K. Kunze, K. Kise, J. Weppner, A. Dengel, P. Lukowicz, and A. Bulling, "In the blink of an eye: Combining head motion and eye blink frequency for activity recognition with google glass," in *Proceedings of the 5th Augmented Human International Conference*, ser. AH '14. New York, NY, USA: ACM, 2014, pp. 15:1–15:4. [Online]. Available: http://doi.acm.org/10.1145/2582051.2582066

[6] B. Zhou, J. Cheng, M. Sundholm, A. Reiss, W. Huang, O. Amft, and P. Lukowicz, "Smart table surface: A novel approach to pervasive dining monitoring," in *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*. IEEE, 2015, pp. 155–162.

[7] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 81–94, 2016.

[8] P. Hevesi, S. Wille, G. Pirkl, N. Wehn, and P. Lukowicz, "Monitoring household activities and user location with a cheap, unobtrusive thermal sensor array," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14. New York, NY, USA: ACM, 2014, pp. 141–145. [Online]. Available: http://doi.acm.org/10.1145/2632048.2636084

[9] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.

[10] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell, "Soundsense: Scalable sound sensing for people-centric applications on mobile phones," in *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '09. New York, NY, USA: ACM, 2009, pp. 165–178. [Online]. Available: http://doi.acm.org/10.1145/1555816.1555834

[11] Y. Yang, B. Guo, Z. Yu, and H. He, "Social activity recognition and recommendation based on mobile sound sensing," in *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*, Dec 2013, pp. 103–110.

[12] N. Eagle and A. S. Pentland, "Social network computing," in *International Conference on Ubiquitous Computing*. Springer, 2003, pp. 289–296.

[13] D. Wyatt, T. Choudhury, J. Bilmes, and J. A. Kitts, "Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 1, pp. 7:1–7:41, Jan. 2011.

[14] T. Nishimura, T. Higuchi, H. Yamaguchi, and T. Higashino, "Detecting smoothness of pedestrian flows by participatory sensing with mobile phones," in *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, ser. ISWC '14. New York, NY, USA: ACM, 2014, pp. 15–18. [Online]. Available: http://doi.acm.org/10.1145/2634317.2642869

[15] J. A. Ward, P. Lukowicz, G. Tröster, and T. E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, October 2006b.

[16] J. A. Ward, P. Lukowicz, G. Pirkl, and P. Hevesi, "Detecting physical collaborations in a group task using Body-Worn microphones and accelerometers," in *13th Workshop on Context and Activity Modeling and Recognition (CoMoRea'17)*, Big Island, USA, Mar. 2017, pp. 268–273.

[17] R. Barea, L. Boquete, M. Mazo, and E. Lopez, "System for assisted mobility using eye movements based on electrooculography," *Trans. on Rehabilitation Engineering*, vol. 10, no. 4, pp. 209–218, December 2002.

[18] M. Wedel, R. Pieters *et al.*, "Eye tracking for visual marketing," *Foundations and Trends® in Marketing*, vol. 1, no. 4, pp. 231–320, 2008.

[19] R. J. Jacob, "Eye tracking in advanced interface design," *Virtual environments and advanced interface design*, pp. 258–288, 1995.

[20] L. A. Granka, T. Joachims, and G. Gay, "Eye-tracking analysis of user behavior in www search," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '04. New York, NY, USA: ACM, 2004, pp. 478–479. [Online]. Available: http://doi.acm.org/10.1145/1008992.1009079

[21] A. Poole and L. J. Ball, "Eye tracking in hci and usability research," *Encyclopedia of human computer interaction*, vol. 1, pp. 211–219, 2006.

[22] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition," in *(Ubicomp) Proceedings of the 11th international conference on Ubiquitous computing*. New York, NY, USA: ACM, 2009, pp. 41–50.

[23] J. Nagel, P. Hevesi, C. Beck, U. Gengenbach, and G. Bretthauer, "Eyelid detection in eye-tracking experiments," *BIOMEDICAL ENGINEERING-BIOMEDIZINISCHE TECHNIK*, vol. 59, pp. S496–S499, 2014.

[24] M. Vidal, A. Bulling, and H. Gellersen, "Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets," in *Proc. UbiComp*. ACM, 2013, pp. 439–448.

[25] M. Dhuliawala, J. Lee, J. Shimizu, A. Bulling, K. Kunze, T. Starner, and W. Woo, "Smooth eye movement interaction using eog glasses," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 2016, pp. 307–311.

[26] Y. Shiga, T. Toyama, Y. Utsumi, K. Kise, and A. Dengel, "Daily activity recognition combining gaze motion and visual features," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 1103–1111.

[27] F. Chollet *et al.*, "Keras," https://github.com/keras-team/keras, 2015, accessed February 15, 2018.

[28] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, "Detectron," https://github.com/facebookresearch/detectron, 2018, accessed February 15, 2018.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[31] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.

[32] A. Fathi, J. K. Hodgins, and J. M. Rehg, "Social interactions: A first-person perspective," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1226–1233.

[33] M. S. Ryoo and L. Matthies, "First-person activity recognition: What are they doing to me?" in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 2730–2737.

[34] H. Pirsiavash and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2847–2854.

[35] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14 Adjunct. New York, NY, USA: ACM, 2014, pp. 1151–1160. [Online]. Available: http://doi.acm.org/10.1145/2638728.2641695

[36] R. Bachu, S. Kopparthi, B. Adapa, and B. Barkana, "Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal," in *American Society for Engineering Education (ASEE) Zone Conference Proceedings*, 2008, pp. 1–7.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[38] J. A. Ward, P. Lukowicz, and H. W. Gellersen, "Performance metrics for activity recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 1, pp. 6:1–6:23, Jan. 2011. [Online]. Available: http://doi.acm.org/10.1145/1889681.1889687

[39] P. Hevesi, "Ward metrics for event based activity recognition evaluation," https://github.com/phev8/ward-metrics, 2017, accessed February 15, 2018.

[40] G. B. H. E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," University of Massachusetts, Amherst, Tech. Rep. UM-CS-2014-003, May 2014.

[41] G. Kalliatakis, "Keras-places," https://github.com/GKalliatakis/Keras-VGG16-places365, 2017, accessed February 15, 2018.

[42] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE transactions on pattern analysis and machine intelligence*, 2017.

Efficacy of Involuntary Deep Breathing by Postural-Respiration Feedback Control System

Samith S. Herath, Kazuki Hayakawa, Osamu Sakai, Ryoma Sekiya, and Shusaku Nomura
Department of Engineering, Nagaoka University of Technology
Nagaoka, Japan
e-mail: s145073@stn.nagaokaut.ac.jp, s143394@stn.nagaokaut.ac.jp, s153406@stn.nagaokaut.ac.jp,
s141049@stn.nagaokaut.ac.jp, nomura@kjs.nagaokaut.ac.jp

Abstract— **This study was undertaken to quantify heart rate dynamics and interactions during sequential use of a postural-respiration feedback regulation control system. The bent shape of the human back makes it easier to exhale air from the lungs, but challenging to inhale. On the other hand, when a human being leans over backwards, it becomes difficult to exhale, but easier to inhale. This simple feature opens up the possibility of regulating human breathing involuntarily through posture control. We developed and tested a postural-respiration feedback regulation control system. The process of the posture control architecture depends on an air chamber placed under the subject's back. For this experiment, subjects had to lie on a bed, and the subjects' respiration cycle was synchronized with the inflation and deflation of the air chamber. We analyzed the beat-to-beat heart rate and the continuous breathing signal data gathered from ten male university students. We investigated the hypothesis that the use of a synchronously controlled postural-respiration feedback control system would yield positive psychophysiological benefits. The heart rate and respiratory dynamics from the deep breathing response and segmented breathing were similar. The results indicate that deep breathing was successfully induced when posture control was precisely synchronized with the subject's own respiration. The implications of these findings for both future research and practice are addressed in a comprehensive discussion.**

Keywords-breathing control; posture feedback; respiration.

I. INTRODUCTION

Breathing is ventilation involving two stages: inspiration and expiration. The physiological background of breathing is the intake of oxygen and removal of carbon dioxide by the lungs. Breathing is essential to support bodily functions of human life. Breathing control has been accepted in modern society around the globe due to its numerous benefits in our daily life. Breathing control methods have become increasingly popular due to the growing significance of a holistic approach to healthcare. In this study, we have developed a new architecture for a breathing control system and investigated the efficacy of the system, which is based on our previous study regarding involuntary deep breathing by posture respiration feedback control system, published at ambient computing applications services and technologies 2017 [1].

A. Psychophysiological Benefits

Breathing techniques are frequently used to change psychophysiological conditions and to improve organ function [2]. Voluntary breathing control improves task performance [3], alleviates anxiety [4]-[6], appropriately controls autonomous nervous system activities such as heart rate, blood pressure [7]-[10], and even immune system functions [11]-[13]. One study reported that the regular practice of slow breathing exercises for three months improves autonomic functions [14], and regular practice of breathing exercises have shown to decrease sympathetic activity and increase vagal tone [15] [16]. Additionally, breathing-related disorders may be dangerous to our daily life, e.g., obstructive sleep apnea is a prominent risk factor for cardiovascular disease [17].

B. Yoga Training and Meditation

There are a variety of methods that have been developed for breathing control such as yoga training and spiritual meditation. Yoga training that uses voluntary breathing control methods is called pranayama [18]. Pranayama involves inhalation, holding, and exhalation, which can be performed either quickly or slowly. Such breathing training can affect oxygen consumption and metabolism [19]. The benefits of pranayama include improving cardiovascular indicators and respiratory functions through increasing vagal tone, decreasing sympathetic discharges, and reducing stress [18] [20]. Additionally, spiritual meditation can be used for breathing control. A slow breathing method is used by Zen monks while performing Zazen meditation [21]. The voluntary breathing control methods used decrease awareness of breathing, create controlled and even breaths, increase mindfulness, and improve respiratory performance.

C. Benefits of Deep Breathing

Numerous studies have suggested that deep breathing exercises have beneficial psychological effects. These studies demonstrate how relaxed, slow, deep breathing significantly reduces negative emotions [22], and yogic deep breathing represses working memory [23]. Additionally, deep breathing techniques have been successfully implemented to enhance academic performance [24]. One study proposed that controlled deep breathing is beneficial for relieving withdrawal symptoms after giving up smoking [25]. Another study stated that the investigations recognized how deep breathing progressively improves ventilation efficiency for oxygen [26]. Deep breathing supports quietening of the mind and body, which can provide psychological and physical benefits.

D. Breathing Control in Clinical Setting

Breathing control has numerous applications in clinical settings. Buteykyo breathing can be one of the most helpful techniques for asthma patients [27]. Using radiotherapy along with deep inspiration with active breathing control may decrease standard tissue irradiation in Hodgkin's disease patients [28]. Training pregnant women in breathing methods is an effective technique for decreasing anxiety, which can influence the duration of the delivery during labor [29]. Another study suggested that the use of deep breathing exercises was helpful after cardiac surgery for first stage patients [30]. Deep breathing exercises and walking have an advantageous effect on the heart rate variability in patients hospitalized for chronic heart failure [31]. In addition to voluntary breathing control, passive ventilation equipment has been developed and has proven to be useful [32] [33].

E. Contribution of the Automatic Nervous System

The autonomic nervous system influences breathing. In the background of the breathing pattern, the autonomic nervous system maintains control of the lungs through ventilation and the gas exchange method, which is the result of the electrical stimulation of nerves and correct physiological stimulus [34]. The autonomic nervous system controls the heart, smooth muscles, and endocrine/exocrine glands. It has both an afferent and an efferent segment [35]. The main branches of the autonomic nervous system are identified as sympathetic and parasympathetic. In sympathetic stimulation, the heart rate increases, whereas in parasympathetic stimulation, the heart rate decreases [36]. The primary characteristic of the autonomic nervous system is that it maintains balance in the body under varying conditions. The hypothalamus can control the body using three different methods. Apart from the ANS, the hypothalamus regulates the endocrine system and a neural system concerned with motivation [37]. It has three main branches: sympathetic, parasympathetic, and central. On a fundamental level, the whole autonomic framework is independent and distantly associated with the other parts of the sensory system.

F. Electrocardiography

Electrocardiogram (ECG) data is used commonly for monitoring the clinical diagnosis of cardiac function. The ECG signal measures the change in electrical potential over the time. Recently, researchers have been able to apply digital signal analysis to the ECG data [38]. Calculating and analyzing ECG signal and time interval of QRS complex existing with in the range between P-Q and Q-T corresponding to different part of the heart behaviors. R complex is the most important point of the QRS complex because several heart beats define R complex as the peek point. Further, beat to beat distance is measured by the milliseconds between each other R complexes as the RR interval. Numerical studies of the cardio system have been using ECG data for more complex diagnostic interpretations [39].

G. Background of Heart Rate Variability

Heart rate variability (HRV) is the measure of beat-to-beat differences in the heart rate influence based on the rate of release of the pacemaker, generally sinus-atrial node. HRV has been used to estimate and investigate cardiac autonomic behaviors in the several fields as clinical, physiological interpretation and process of ECG. HRV can be measured in a wide range of time intervals, such as for a short 2-minute interval or an extended period of 24 hours. HRV can be used to analyze Cardiac arrhythmia conditions by using the time domain, frequency domain, and nonlinear methods [40].

Time domain variable calculation involves the mean peak-to-peak interval (RR interval). The standard deviation of the RR intervals, or SDRR, is frequently used as an index of sympathetic nervous system activity.

The frequency-domain measure is obtained using the Fast-Fourier transform method. The power of the heart rate variation includes both high frequency (HF) and low frequency (LF) ranges. The HF commonly range from 0.15 to 0.40 Hz, while the LF commonly range from 0.05 to 0.15 Hz. The heart rate variation in the HF segment is considered to be an index of parasympathetic nervous system activity [41]. The LF segment reflects both sympathetic and parasympathetic nervous system activity, so the rate LF/HF is frequently introduced as an index of sympathetic–vagal balance [42].

H. Hypothesis and Objective of this Study

As stated earlier, breathing control can bring positive psychological and physiological effects. However, there are some limitations in both voluntary and passive breathing control methods. First of all, voluntary breathing control requires significant effort and concentration by the patient, and it is difficult to continue over a long period of time, especially for the elderly. Additionally, passive ventilation, such as continuous positive airway pressure (CPAP) [32] and adaptive servo-ventilation (ASV) [33] require patients to be fitted with a nasal cannula or respirator, so it is not suitable for daily use at home.

Therefore, in this study, we propose an alternative method for semi-passive breathing control via posture control. The idea is based on the bent shape of the human backbone making it easier to exhale but challenging to inhale. However, when a human being is leaning backwards, it becomes difficult to exhale, but easier to inhale. This simple feature opens the possibility to regulate human breathing involuntarily through posture control. Based on this simple idea, we developed a postural-respiration feedback regulation control system and tested the efficacy of the proposed architecture.

We developed an architecture that forces the user to change their posture synchronously with their own breathing cycle. In the next section, the architecture of the developed system and the procedure of the experiment to test the efficacy of the system are described. The following sections describe the results of the experiment, discussion, and conclusion.

II. Method

This study aimed to develop an involuntary postural-respiration feedback regulation control system. We concentrated on the fact, that changing posture affects respiration. This physical condition of the bent shape of the human back makes it easier to exhale, but challenging to inhale. However, when a person is leaning backwards, it is difficult to exhale but easier to inhale. The idea of posture affecting respiration opens the possibility for breathing control by changing posture. To verify this idea, we developed a postural-respiration feedback regulation control system. Figure 1 shows the architecture of the system.

A. Materials

In this system a silicon air chamber is placed beneath the subject's back. The air chamber is synchronized to deflate/inflate according to the subject's own respiration cycle. For example, when a subject starts inhaling/exhaling, the air-chamber starts to inflate/deflate to assist the user in inhalation/exhalation of the air into/from the lungs.

The basic architecture of this system is shared with our previous studies [43]. A thermistor sensor detects the subject's respiration pattern in a time series. This analogue for respiration is amplified and digitized by analog-to-digital converter (NI USB-6008 DAQ, National Instruments Co., USA), and transmitted to the feedback regulator. The feedback regulator, developed with a visual programming language (LabVIEW, National Instruments Co., USA), controls the air flow to the chamber using a pair of air pumps (YP-20A, Yasunaga Air Pump Inc., Japan) and a switching circuit to regulate the volume of the air chamber.

With this adaptive architecture, we assume that such an adaptive intervention to the posture would make the subject's breathing deeper and longer, and the subject would be more relaxed in body and mind. To test our hypothesis, we conducted an experiment with three different respiratory intervention conditions: 1) the air-chamber inflates and deflates synchronously with subject's inhalation and exhalation (called "SNC" hereafter); 2) the air-chamber asynchronously inflates and deflates with breathing: inflating when the subject exhales and inflating when the subject inhales (called "ASN" hereafter); and 3) the air-chamber inflates and deflates with a constant rhythm (called "CST" hereafter). In SNC and ASN conditions, the inflation and

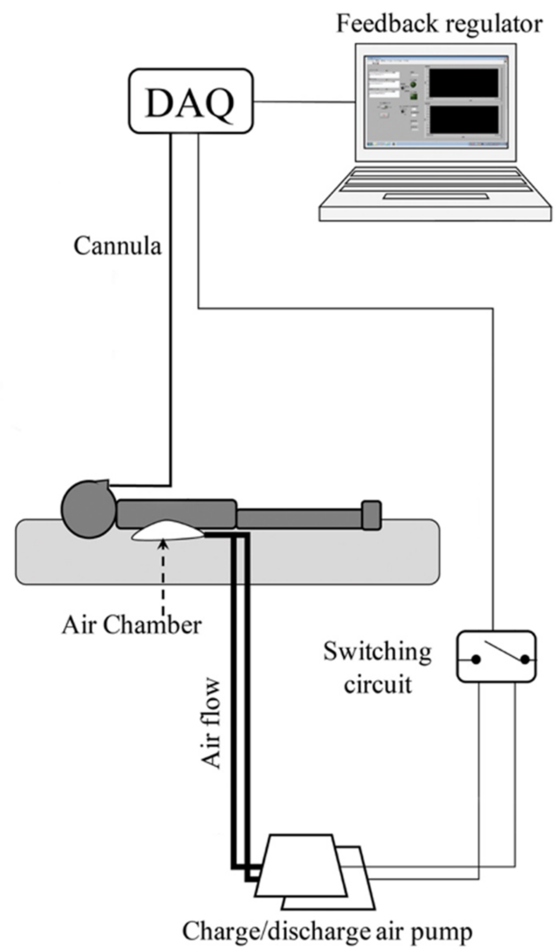


Figure 1. Schematic diagram of the posture-respiration feedback regulation control system.

deflation cycle of the air-chamber is precisely determined by the subject's respiration steps, whereas in the CST condition, the cycle of the air chamber is set to the same duration as each subject's mean deep breathing cycle, which has been determined in advance.

Figure 2 shows an example of the respiration pattern and the movement of the air chamber regulated in the three different conditions. In the SNC condition, when the subject inhales, the air chamber continues to inflate and when the subject exhales the air chamber continues to deflate. Accordingly, the air chamber inflates and deflates synchronously with the subject's respiration pattern.

Figure 3 shows how the posture of a subject changes with the inflation and deflation of the air chamber placed under the subject's back. It should be noted that the movement of the air chamber and following change in the posture in this image is exaggerated to illustrate how the system works. The actual movement of the air chamber and the change in the posture, was smaller than in this image in order to not place any burden on the subjects.

Compared with SNC, the ASN condition shows the opposite pattern as the air chamber asynchronously inflates and deflates with subject's inhalation and exhalation. In the CST condition, the air chamber inflates and deflates with a constant rhythm determined by each subject's mean deep breathing cycle.

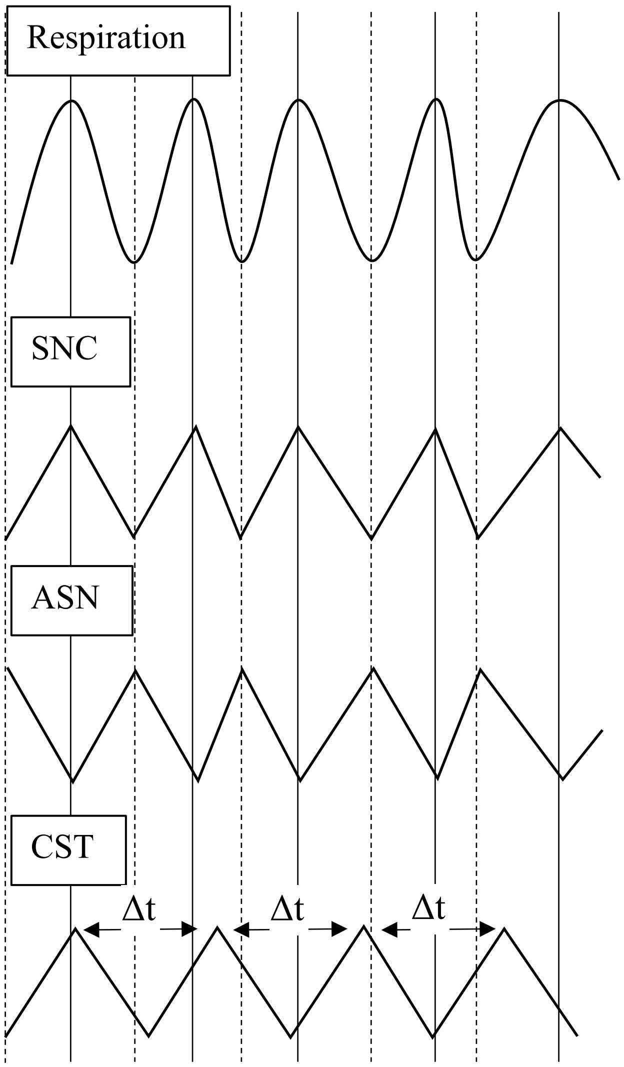


Figure 2 Patten of reparation and other conditions

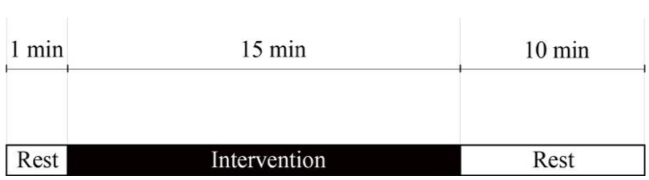


Figure 4. Experiment schedule

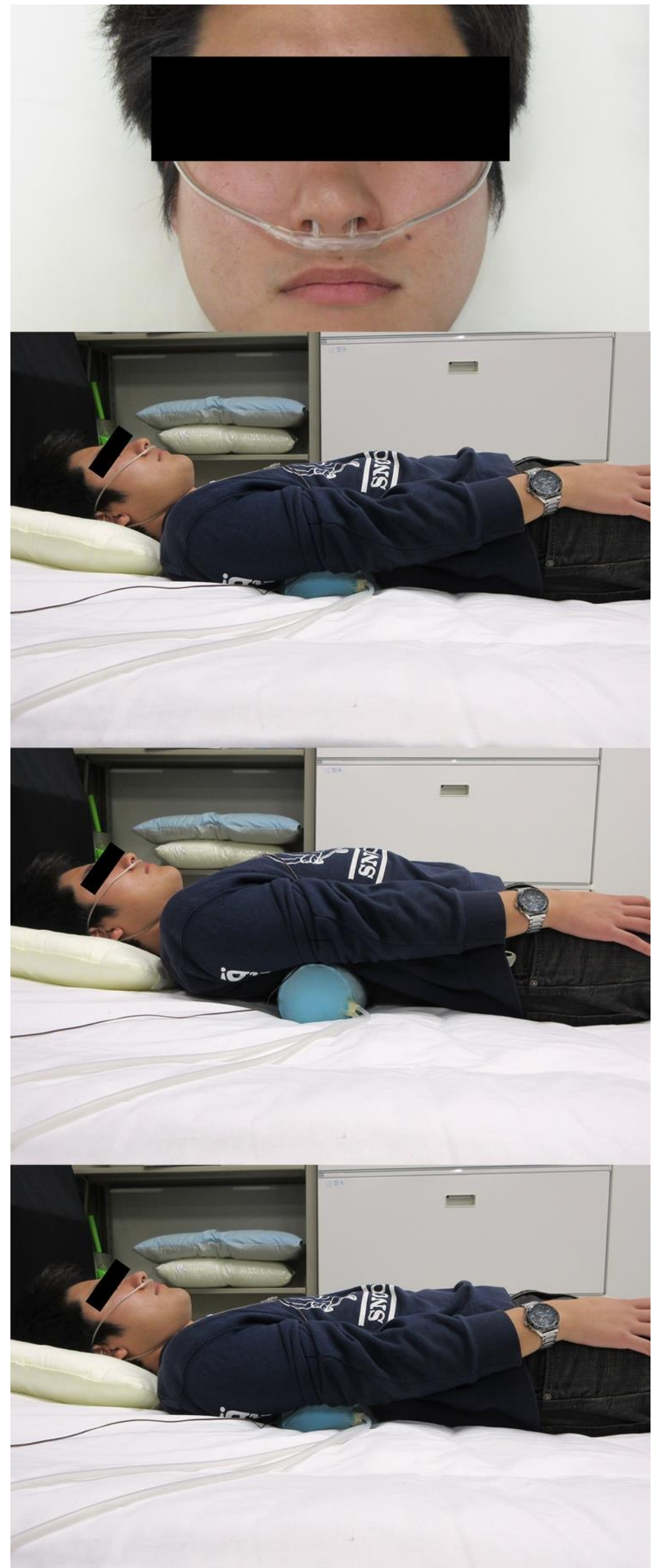


Figure 3. Postural-respiration feedback control system

B. Experiment

Subjects (n = 10) were male university students (age 22-24) and they voluntarily participated in this experiment. All subjects were rated for their health condition, and they had no heart or breathing related health problems. The study was conducted following ethical principles and informed consent was obtained from all subjects. The study was approved by the ethics committee of the Nagaoka University of Technology.

Figure 4 shows the experiment protocol. The experiment consisted of a 1-minute initial rest period, followed by 15 minutes of the intervention period, and another 10-minute rest period (recovery period). The total experiment time was 26 minutes. The postural-respiration feedback control system functioned only during the intervention period. A thermistor sensor measured respiration of the subjects. The heartbeat signal was measured by an electrocardiogram (ECG) with a bio-signal amplifier system (MP150, BIOPAC Systems, Inc., USA) to evaluate the cardio-physiological functioning by the postural-respiration intervention. The subject was instructed to lie on the bed with the rubber air chamber placed under the subject's back. In this experiment, the heart rate (HR) and the high frequency (HF) component of its variation were measured using the ECG data. The experiment was conducted in a within-subject design such that each participant went through all three conditions in one trial per day in a randomized order.

III. Results

The postural-respiration feedback regulation control system (Figures 2 and 3) produced results in the dynamic change in the beat-to-beat respiration interval (RI) and respiration amplitude (RA) of the respiration during the intervention period. Figures 5 through 8 show standardized values to compensate for the large variation among the individuals (Z-score). Figure 5 shows that RI response and the segment breathing induced significantly longer respiration (RI: $p < 0.01$) in the SNC condition. For the RA, Figure 6 shows a deeper response (RA: $p < 0.01$) in SNC condition. This means that in the intervention period the subject's respiration effectively synchronized with the air-chamber. Compared to the recovery period, the interval and amplitude were significantly changed during the intervention duration.

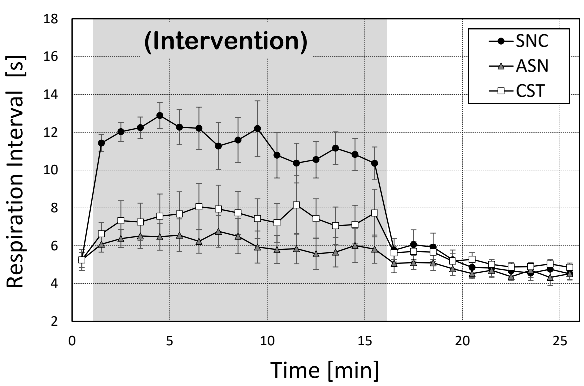


Figure 5. Change in the respiration interval (mean±S.E.M).

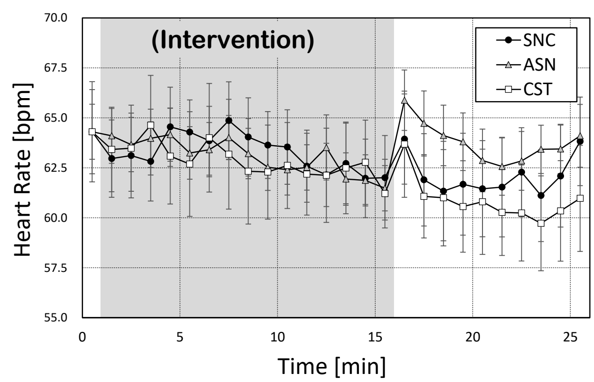


Figure 7. Change in the heart rate (mean±S.E.M).

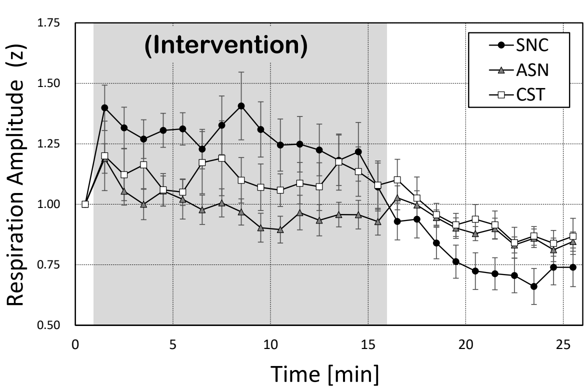


Figure 6. Change in the respiration amplitude (mean±S.E.M).

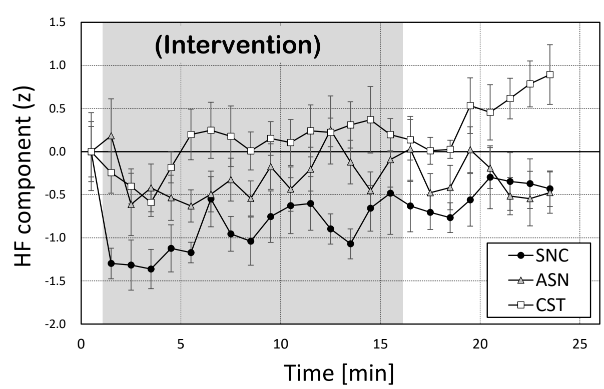


Figure 8. Change in the HF component (mean±S.E.M).

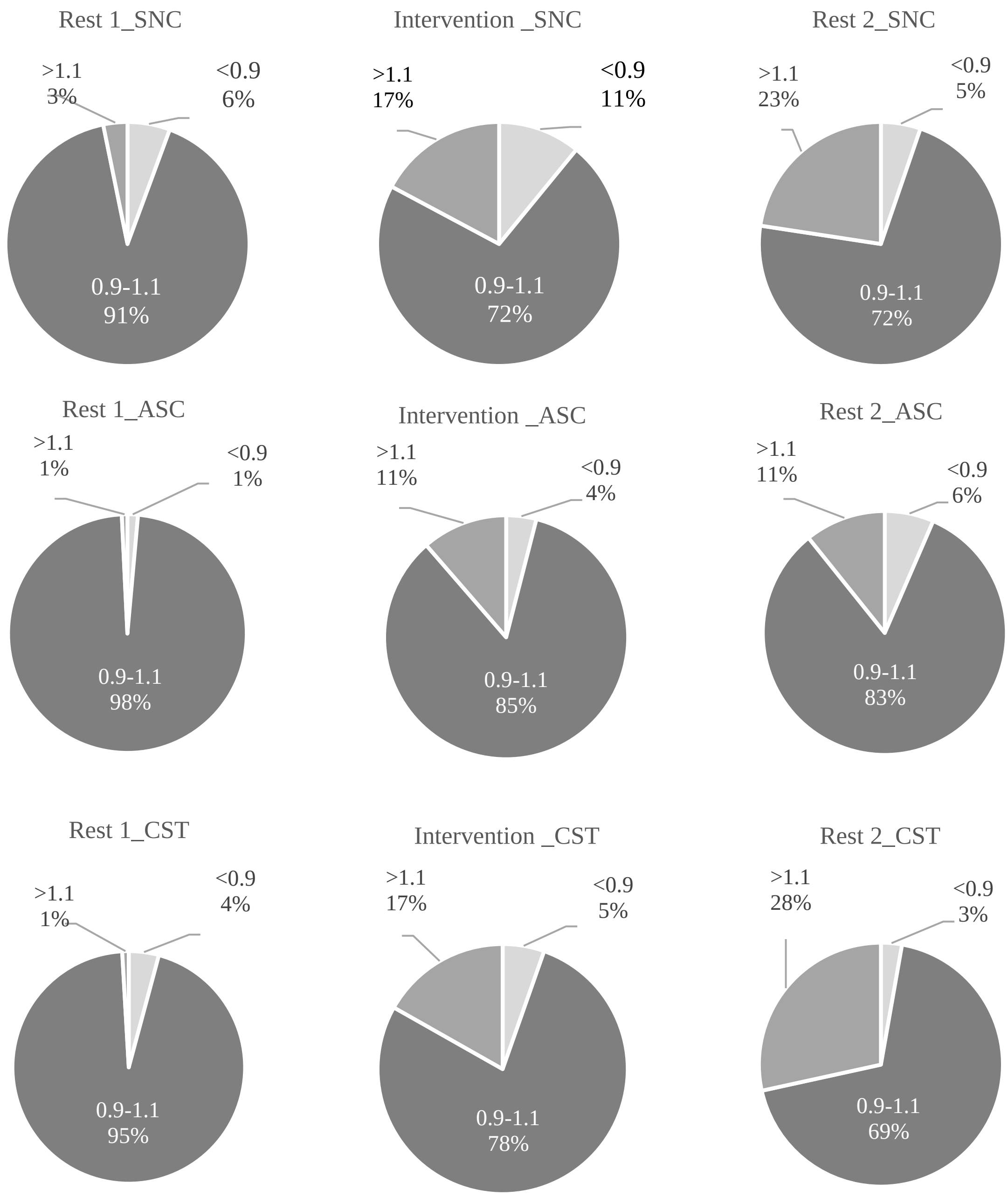


Figure 9. The ratio of RRI change.

As shown in Figures 5 and 6, in ASN condition the subject's breathing was remarkably restricted in intervention period. While, in CST condition, the change in respiration is in between SNC and ASN as shown in the graphs. The frequency analysis of the heart rate variability was used to evaluate the postural-respiration feedback system. Figures 7 and 8 show the results of the change in the heart rate (HR) and high frequency (HF: 0.15 - 0.40 Hz) component of the HRV. As seen in Figure 7, HR did not significantly change in any of the three conditions ($p > 0.05$) during the intervention period. However, HR in the recovery period in ASN was significantly higher than that of other conditions ($p < 0.01$), and it reached a higher value than the initial rest period (overshoot). With regard to the HF component in the SNC condition as shown in Figure 8, it was the lowest during the intervention period and recovery period ($p < 0.01$).

Cardiac vagal activity was measured by the ECG recording, and the HRV was calculated from this data. The

HF component of the HRV in SNC condition has an instantaneous R-R interval (RRI) result shown in Figure 9. The R-R interval data has been analyzed to identify vagal activation with inclusion of feedback SNC, non-feedback ASN, and control CST during the experiment period. The average values of the RRI data during the Rest 1, Intervention, and Rest 2 periods were calculated. The ratio of the x is obtained by dividing the value of RRI when the average value is 0. The RRI results are shown by pie charts in Figure 9, either $x < 0.9$, $0.9 \leq x \leq 1.1$, or $1.1 < x$ values. The results of the statistical test are show in Table I. Heart rate variability and Rest 2 periods tended to increase in all intervention conditions in the feedback period, and there was a significant difference between SNC and ASN ($0.9 \leq x \leq 1.1$) and between ASN and CST ($0.9 < x < 1.1$) in the Rest 2 period the results with a p value $p < 0.05$. These results show vagal activity of the HF component. The results from the rest periods without movement of the chamber also show cardiac vagal activation.

Table I. RRI RATIO AMONGST CONDITIONS

SNC vs ASN	SNC vs CST	ASN vs CST	RRI	*p* value
Intervention period				
0.061	0.114	0.362	$x < 0.9$	$p < 0.05$
0.013	0.428	0.243	$0.9 \leq x \leq 1.1$	
0.276	0.957	0.360	$x > 1.1$	
Rest 2 period				
0.507	0.224	0.020	$x < 0.9$	$p < 0.05$
0.221	0.683	0.113	$0.9 \leq x \leq 1.1$	
0.187	0.541	0.042	$x > 1.1$	

The result of paired *t* – test, RR interval vagal activation intervention period and rest 2 period.

IV. DISCUSSION

In this study, the impact of the postural-respiration feedback control system on respiration and autonomous cardiac system function was investigated. The hypothesis examined the efficacy of the proposed architecture accompanied by three different respiratory intervention conditions. Our results were as expected in SNC condition, meaning that the subject's respiration (inhalation/exhalation) synchronized with air chamber inflation/deflation successfully. The results of Figures 5 and 6 show (RI: $p < 0.01$) that respiration was made significantly longer, (RA: $p < 0.01$) and significantly deeper. This result implies that the developed posture control system can induce deep breathing unconsciously.

However, in ASN condition where the posture regulation was administered asynchronously with respiration (i.e., the phase delay is 180 degrees), there was no change in the respiration interval and amplitude. Among the components, moreover, in the recovery period (Rest 2), an overshooting of HR was observed in ASN condition. HR overshooting is frequently observed in the recovery period after mild exercise [44]. It occurs to meet the requirement of balancing the greater oxygen cost of fat catabolism during the early recovery period. It may be that if the HR overshoot seen in this study shares part of the same background with that of after exercise, it might reflect a lack of oxygen during the intervention period. Despite the large effort of doing respiration, whether it be intentional or unconscious, ASN condition made breathing difficult, and this phenomenon misled the body into believing that the concentration of air oxygen was "low", resulting in a higher requirement for oxygen after intervention (during "normal" circumstances).

In CST, the impact on respiration and cardiac autonomous system were marginal compared to SNC and ASN, and the breathing seemed to be gradually entrained with the constant rhythm. The cycle of the air chamber movement in CST was fixed at the same duration as each subject's mean deep breathing cycle, so subjects might modulate their breathing cycle voluntarily to their deep breathing. However, the interval and amplitude of the respiration in SNC reached far beyond CST. In this sense, our developed posture-respiration intervention architecture has demonstrated a prominent impact on the respiration.

However, HF in SNC was strongly suppressed as shown in Figure 8. The HF component of the heart rate variability has been frequently taken as an index of cardiac parasympathetic nervous system activation [45]. However, as demonstrated in Figure 9 and Table I, the change in RR-interval, i.e., the variation of the instant heart rate, was prominent in SNC. Therefore, it is unable to interpret the result of our study as the SNC condition suppressing parasympathetic nervous activity.

The physiological background of the above mentioned phenomenon is the relationship between deep breathing and cardiac behavior. The central school of thought in physiology discusses on how respiratory sinus arrhythmia (RSA) and cardiac vagal tone contribute to the heart rate. In the normal condition of RSA in the respiratory process, the HR increases during inspiration and decreases during expiration; this is called RSA. Sinus arrhythmia is a normal physiological phenomenon most typically seen in young, healthy people. RSA can be utilized as the key to understanding the cardiac vagal tone and contributes to heart rate variability [46]. RSA has been shown to significantly contribute to the rhythmic waxing and waning of cardiac vagal efferent effects upon the sinoatrial node and, therefore, the heart rate [47]. The respiratory rate and tidal volume produce independent and interactive effects [48]. This physiological background contributes to the understanding of the breathing process as a gas exchange mechanism. Numerous studies have documented how gas exchange productivity improves with deep breathing and increased mean heart rate, but in this process it is unconnected to the RSA contribution. Among the studies mentioned, HR was remarkably similar to RSA with slow and deep breaths [49].

The result of the posture control system we developed showed that deep breathing was successful at producing the desired effects. If the participant were to breathe in deeply, the heart rate would increase. That is because the heart is making the most of a fresh supply of oxygen to the lungs and sending a ration of blood through the lungs to receive the oxygen. When breathing out, it is more energy-efficient for the heart not to pump as fast, because there is less oxygen available in the lungs. This reaction, controlled by the brain, detects a slight change in the gas composition of the blood. This is detected and sent by the vagus nerve to the sinoatrial node. That node cell is at the top of the heart. The node works as a heart pacemaker meaning that the heart rate is changed by a deep breath. In this study, the heart rate condition did not significantly change during the conditions ($p > 0.05$) in the intervention period, but it seems that a deep breath with gas exchange changes the cardiac mechanism by means of RSA. From the above, it can be concluded that deep breathing gives rise to RSA that is more pronounced than standard RSA. In the HF analysis using the HRV method, the response is about 0.3 Hz (HF). Additionally, it seems that the HF component of the SNC condition has decreased. Frequency domain analysis is the most accurate method of discovering correlations between the HF and LF components and autonomic activation. HRV measures can be associated with several cardiac functions with the assumption of sympathetic or parasympathetic activation. Normally, the LF condition is associated with sympathetic activation. Generally, the HF component is associated with vagal modulation and respiration because HF is dependent on the respiration pattern.

In our study, the high-frequency component of the RR interval results indicated cardiac vagal activation in the HF component. These results are shown in Figure 9. The changes in RR intervals describe the cardiac vagal activation during the intervention period and the rest period. These results are a valuable part of this study because the vagal activation response contributes both physiological [50] and psychological benefits. The vagus nerves also control the central part of the parasympathetic nervous system. Their influence on breathing, digestive function, and heart rate has an impact on mental health conditions.

A. Limitation

This study was conducted in a within-subject manner with randomized order which can free from group effect. However, the small sample size ($n = 10$) with homogeneous population limits the generalization of the interpretation of the results. The strength of posture control, i.e., the air pressure in the air chamber, was not varied regardless of the attribute of each subjects, which can also form the limitation of this study

B. Implications and Future Research

Using this architecture, inspired by involuntary deep breathing, can create psychophysiological benefits. After the development and modification of a clinical instrument, this posture control system can be used for clinical treatment for patients with respiration problems. Further, in future investigations, psychological and physiological changes can be evaluated using the different types of parameters with a more significant number of subjects. In the future, it is expected that this posture control system can be developed and improved as a stress-reducing mechanism with an involuntary breathing control and as a health product for the global market. Future research using this architecture will improve psychophysiological benefits through an involuntary deep breathing and inspiration.

V. Conclusion

In summary, we tested the efficacy of our postural-control respiration intervention system in the manner of real-time feedback regulation. The results indicate that deep breathing is successfully induced when the posture control is precisely synchronized with the user's own respiration. Regarding the impact on cardiac function, HR overshoot was observed as the after effect of asynchronous regulation. We found this from the results of cardiac vagal activation. Further physiological study promises to reveal the background mechanism of the impact of our system.

Acknowledgment

This work was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 15H02767 and 16K12510.

References

[1] S. Nomura, K. Hayakawa, O. Sakai, R. Sekiya, and S.S. Herath, "Involuntary "Deep Breathing" by Posture-Respiration Feedback Control System," *Proc. The Seventh International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT2017)*, Barcelona (Spain), 2017, pp. 1-4.

[2] T. Ritzand and W.T. Roth, "Behavioral interventions in asthma: Breathing training." Behavior Modification, vol. 27, no. 5, 2003, pp. 710-730.

[3] A. Solanky, "Respiration biofeedback assisted controlled breathing training to enhance shooting performance," Br. J. Sports. Med, vol. 44, 2010, pp. i27-i28.

[4] A. P. Sutarto, M. N. Wahab, and N. M. Zin, "Resonant breathing biofeedback training for stress reduction among manufacturing operators," Int. J. Occup. Saf. Ergon, vol. 14, no. 4, 2012, pp. 549-561.

[5] W. Wu, Y. Gil, and J. Lee, "Combination of wearable multi-biosensor platform and resonance frequency training for stress management of the unemployed population," Sensors (Basel), vol. 12, no. 10, 2012, pp. 13225-13248.

[6] Y. H. Su, J.J. Luh, H.I. Chen, C.C. Lin, M.J. Liao, and H.S. Chen, "Effects of using relaxation breathing training to reduce music performance anxiety in 3rd to 6th graders," Medical Problems of Performing Artists, vol. 25, no. 2, 2010, pp. 82-86.

[7] L. Guan-Zheng, H. Bang-Yu, and W. Lei, "A wearable respiratory biofeedback system based on generalized body sensor network," Telemedicine and e-Health, vol. 17, no. 5, 2011, pp. 348-357.

[8] P. M. Lehrer, E. Vaschillo, and B. Vaschillo, "Resonant frequency biofeedback training to increase cardiac variability:

rationale and manual for training," Appl. Psychophysiol Biofeedback, vol. 25, no. 3, 2000, pp. 177-191.

[9] Q. A. Morarend, M. L. Spector, D. V. Dawson, S. H. Clark, and D. C. Holmes, "The use of a respiratory rate biofeedback device to reduce dental anxiety: an exploratory investigation," Applied Psychophysiol Biofeedback, vol. 36, no. 2, 2011, pp. 63-70.

[10] S. Z. Wang, S. Li, X.Y. Xu, G.P. Lin, L. Shao, Y. Zhao, and T.H. Wang, "Effect of slow abdominal breathing combined with biofeedback on blood pressure and heart rate variability in prehypertension," The Journal of Alternative and Complementary Medicine, vol. 16, no. 10, 2010, pp. 1039-1045.

[11] M.S. Rider, J. Achterberg, G.F.Lawlis, A.Goven, R.Toledo, and J.R. Butler, "Effect of immune system imagery on secretory IgA," Biofeedback Self Regulation, vol. 15, no. 4, 1990, pp. 317-333.

[12] R. G. Green and M. J. Green, "Relaxation increases salivary immunoglobulin A," Psychological Reports, vol. 61, no. 2, 1987, pp. 623-629.

[13] M. L. Jasnoski and J. Kugler, "Relaxation, imagery, and neuroimmunomodulation," ANNALS of The New York Academy of Sciences, vol. 496, 1987, pp. 722-730.

[14] G.K. Pal and S. Velkumary,"Effect of short-term practice of breathing exercises on autonomic functions in normal human volunteers," Indian Journal of Medical Research, vol. 120, no. 2, 2004, pp. 115.

[15] S.Telles, R. Nagarathna, and H. R. Nagendra, "Breathing through a particular nostril can alter metabolism and autonomic activities," Indian Journal of Physiology and Pharmacology, vol. 38, 1994, pp. 133-133.

[16] A.Jr. Stancák, M. Kuna, S. Vishnudevananda, and C. Dostalek, "Kapalabhati--yogic cleansing exercise. I. Cardiovascular and respiratory changes," Homeostasis in health and disease: international journal devoted to integrative brain functions and homeostatic systems, vol. 33, no. 3, 1991, pp. 126-134.

[17] H.K. Yaggi, J. Concato, W.N. Kernan, J.H. Lichtman, L.M. Brass, and V. Mohsenin, "Obstructive sleep apnea as a risk factor for stroke and death," The New England Journal of Medicine, vol. 353, no. 19, 2005, pp. 2034-2041.

[18] V.K. Sharma, M. Trakroo, V. Subramaniam, M. Rajajeyakumar, A. B. Bhavanani, and A. Sahai, "Effect of fast and slow pranayama on perceived stress and cardiovascular parameters in young health-care students," International journal of yoga, vol. 6, no. 2, 2013, pp. 104-110.

[19] J. Ravinder, J.W. Edry, V.A. Barnes, and V. Jerath, "Physiology of long pranayamic breathing: Neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system," vol. 67, no. 3, 2006, pp. 566-571.

[20] R. Bhargava, M.G. Gogate, and J.F. Mascarenhas, "Autonomic responses to breath holding and its variations following pranayama," Indian J Physiol Pharmacol, vol. 32, no. 4, 1988, pp. 257-64.

[21] Lehrer, Paul, Y.Sasaki, and Y. Saito. "Zazen and cardiac variability," Psychosomatic medicine, vol. 61, no. 6, 1999, pp. 812-821.

[22] V.B, Magerl, W. Magerl, U.Kern, J.Haas, G. Hajak, and P. Eichhammer, "The effect of deep and slow breathing on pain perception, autonomic activity, and mood processing-an experimental study," Pain Medicine, vol. 13, No. 2, 2012, pp. 215-228.

[23] N. Gothe, M.B. Pontifex, C. Hillman, and E. McAuley, "The acute effects of yoga on executive function," Journal of physical activity and health, vol. 10, no. 4, 2013, pp. 488-495.

[24] G.Paul, B. Elam, and S.J. Verhulst, "A longitudinal study of students' perceptions of using deep breathing meditation to reduce testing stresses," Teaching and learning in medicine, vol. 19 no. 3, 2007, pp. 287-292.

[25] F. J. McClernon, E.C. Westman, and J.E. Rose, "The effects of controlled deep breathing on smoking withdrawal symptoms in dependent smokers," Addictive Behaviors, vol. 29, no. 4, 2004, pp. 765-772.

[26] G. Bilo, M. Revera, M. Bussotti, D. Bonacina, K. Styczkiewicz, G. Caldara, A. Giglio, A. Faini, A. Giuliano, C. Lombardi, K. K. Jaszcz, G. Mancia, P. Agostoni, and G. Parati, "Effects of slow deep breathing at high altitude on oxygen saturation, pulmonary and systemic hemodynamics," PloS one 7, no. 11, 2012, e49074.

[27] A. J. Opat, M. M. Cohen, M. J. Bailey, and M. J. Abramson, "A clinical trial of the Buteyko Breathing Technique in asthma as taught by a video," Journal of Asthma, vol. 37, no. 7, 2000, pp. 557-564.

[28] J.S. Stromberg, M.B. Sharpe, L.H. Kim, V.R .Kini, D.A. Jaffray, A.A. Martinez, and J.W. Wong, "Active breathing control (ABC) for Hodgkin's disease: reduction in normal tissue irradiation with deep inspiration and implications for treatment," International Journal of Radiation Oncology• Biology• Physics, vol. 48, no. 3, 2000, pp. 797-806.

[29] S. Cicek and F. Basar. "The effects of breathing techniques training on the duration of labor and anxiety levels of pregnant women," Complementary therapies in clinical practice, vol. 29, 2017, pp. 213-219.

[30] C. Urell, M. Emtner, H. Hedenström, A. Tenling, M. Breidenskog, and E. Westerdahl. "Deep breathing exercises with positive expiratory pressure at a higher rate improve oxygenation in the early period after cardiac surgery-a randomised controlled trial," European Journal of Cardio-thoracic Surgery, vol. 40, no. 1, 2011, pp. 162-167.

[31] F.C. Caruso, R.A.R.G. Mendes, M.S. Reis, V. Papa, and A. B. Silva. "Heart rate autonomic responses during deep breathing and walking in hospitalised patients with chronic heart failure," Disability and rehabilitation, vol. 33, no. 9, 2011, pp. 751-757.

[32] J. M. Marin, S. J. Carrizo, E. Vicente, and A. G. N. Agusti, "Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study," The Lancet, vol. 365, 2005, pp. 1046-1053.

[33] P. C. Hastings, A. Vazir, G.E. Meadows, M. Dayer, P.A.P. Wilson, H.F. Mclntyre, M.J. Morrell, M.R. Cowie, and A.K. Simonds, "Adaptive servo-ventilation in heart failure patients with sleep apnea: A real world study," International Journal of Cardiology, vol. 139, 2010, pp. 17-24.

[34] J. G. Widdicombe and G.M. Sterling, "The autonomic nervous system and breathing," Archives of internal medicine, vol. 126, no. 2, 1970, pp. 311-329.

[35] K. Nishijo, H. Mori, K. Yosikawa, and K. Yazawa. "Decreased heart rate by acupuncture stimulation in humans via facilitation of cardiac vagal activity and suppression of cardiac sympathetic nerve," Neuroscience letters, vol. 227, no. 3, 1997, pp. 165-168.

[36] I.M. Korr, "The spinal cord as organizer of disease processes:(ll) The peripheral autonomic nervous system," The collected papers, vol. 2, 1979, pp. 45-53.

[37] D. G. Amaral, "Integration of sensory and motor function: the association areas of the cerebral cortex and the cognitive capabilities of the brain," Principles of neural science, 2000.

[38] D. P.Golden, R. A. Wolthuis, and G. W. Hoffler. "A spectral analysis of the normal resting electrocardiogram," IEEE Transactions on Biomedical Engineering, vol. 20, no. 5, 1973, pp. 366-372.

[39] L.De Lathauwer, B. De Moor, and J. Vandewalle. "Fetal electrocardiogram extraction by blind source subspace

separation," IEEE transactions on biomedical engineering, vol. 47, no. 5, 2000, pp. 567-572.

[40] M.T.L. Rovere, J.T.B. Jr, F.I. Marcus, A. Mortara, and P.J. Schwartz, ATRAMI (Autonomic Tone and Reflexes After Myocardial Infarction) Investigators, "Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction," The Lancet, vol. 351, no. 9101, 1998, pp. 478-484.

[41] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, Philippe Coumel et al, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," European heart journal, vol. 17, no. 3, 1996, pp. 354-381.

[42] S. C. Malpas and G. L. Purdie, "Circadian variation of heart rate variability," *Cardiovascular research,* vol. 24, no. 3, 1990, pp. 210-213.

[43] S. Nomura and A. Kusumi, "Respiration-Posture Feedback System for Breathing Control," Proc. The Third International Conference on Ambient Computing, Applications, Services and Technologies, AMBIENT2013, pp. 17-22.

[44] D. Joshi, A. Shiwalkar, M.R. Cross, S.K. Sharma, A. Vachhani, and C. Dutt, "Continuous, non-invasive measurement of the haemodynamic response to submaximal exercise in patients with diabetes mellitus: evidence of impaired cardiac reserve and peripheral vascular response." Heart. vol. 96, no. 1, 2010, pp. 36-41.

[45] B. Pomeranz, R. J. Macaulay, M. A. Caudill, I. Kutz, D. Adam, D. Gordon, K. M. Kilborn, A. C. Barger, D. C. Shannon, and R. J. Cohen, "Assessment of autonomic functions in human by heart rate spectral analysis," Am. J. Physiol., vol. 248, no. 1, 1985, pp. 151-153.

[46] P.Grossman and E.W Taylor "Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions," Biol Psychol, vol. 74 no. 2, 2007, pp. 263–285.

[47] D.L. Eckberg, "The human respiratory gate," The Journal of Physiology, vol. 548, no. 2, 2003, pp. 339 -352.

[48] J.A. Hirsch and B. Bishop, "Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate," American Journal of Physiology-Heart and Circulatory Physiology, vol. 241, no. 4, 1981, pp. 620-629.

[49] A. Ben-Tal, S.S. Shamailov, and J.F.R.Paton, "Evaluating the physiological significance of respiratory sinus arrhythmia: looking beyond ventilation–perfusion efficiency," The Journal of physiology, vol. 590, no. 8, 2012, pp. 1989-2008.

[50] P.W.Stephen, J, A. Doussard-Roosevelt, and A. K. Maiti. "Vagal tone and the physiological regulation of emotion," Monographs of the society for research in child development, vol. 59, no. 2-3, 1994, pp. 167-186.

www.iariajournals.org

International Journal On Advances in Intelligent Systems
issn: 1942-2679

International Journal On Advances in Internet Technology
issn: 1942-2652

International Journal On Advances in Life Sciences
issn: 1942-2660

International Journal On Advances in Networks and Services
issn: 1942-2644

International Journal On Advances in Security
issn: 1942-2636

International Journal On Advances in Software
issn: 1942-2628

International Journal On Advances in Systems and Measurements
issn: 1942-261x

International Journal On Advances in Telecommunications
issn: 1942-2601