Fuzzy Model-based Design of a Transparent Controller for a Time Delayed Bilateral Teleoperation System Through State Convergence

Umar Farooq[1,3], Jason Gu[1], Mohamed E. El-Hawary[1], Valentina E. Balas[2], Muhammad Usman Asad[4], Ghulam Abbas[4]

[1]Department of Electrical and Computer Engineering Dalhousie University, Halifax, N.S. B3H 4R2, Canada

[2]Department of Automatics and Applied Software, University of "Aurel Vlaicu" Arad, Romania

[3]Department of Electrical Engineering, University of The Punjab, Quaid-e-Azam Campus, Lahore, 54590 Pakistan

[4]Department of Electrical Engineering, The University of Lahore, Pakistan

umar.farooq@dal.ca, jason.gu@dal.ca, elhawary@dal.ca, valentina.balas@uav.ro, usman.asad@ee.uol.edu.pk, ghulam.abbas@ee.uol.edu.pk

Abstract: Transparency optimized state convergence scheme is a simple, elegant and easy to design bilateral control algorithm for n^{th}-order linear teleoperation systems modeled on state space. It requires the solution of 3n+2 design equations to determine the control gains when the time delay in the communication channel is small. The controller thus obtained allows the slave system to follow the master system's motion in a desired dynamic way while providing the human operator with a high degree of environment's feeling during the steady state operation. This paper employs the transparency optimized state convergence scheme to design a bilateral controller for a class of nonlinear teleoperation systems which can be approximated by Takagi-Sugeno (TS) fuzzy models. To close the feedback loop around master and slave systems, a suitable parallel distributed compensation (PDC) type TS fuzzy controller is selected which allows to use the state convergence procedure in a true sense. In this way, all the benefits of the state convergence scheme have been kept intact while extending its applicability to nonlinear teleoperation systems at the same time. Further, the proposed fuzzy state convergence controller is more general as the existing linear state convergence controller can be derived from it. MATLAB simulations on a one-degree-of-freedom (DoF) nonlinear teleoperation system are included to show the efficacy of the proposed fuzzy transparent state convergence controller.

Keywords: state convergence; bilateral teleoperation systems; TS fuzzy models; MATLAB/Simulink

1 Introduction

Teleoperation refers to the distant control of a process. The main components of a teleoperation system are the human operator, master system, communication channel, slave system and the environment. The system works under the control of a human operator who generates the desired motion for the slave system by driving the master system and these motion commands are then sent over the communication channel to the slave system which executes the required task on the remote environment. This system is known as unilateral. However, if the slave system also sends a task-related (usually force) feedback to the master system then the teleoperation system is said to be bilateraly controlled. The popularity of such bilateral systems can be judged from their deployment in diverse processes ranging from miniaturized surgical tasks to large scale industrial systems [1].

The control design of bilateral systems is challenging mainly due to the presence of time delay in the communication channel and the performance is further deteriorated due to the other form of systems' uncertainties. The main objectives for the bilateral controller are to guarantee the stability in the presence of time delays and to provide a high degree of transparency (position and force tracking) [2]-[4]. Passivity based control algorithm has been widely used in bilateral systems owing to its robustness to time delays [5]-[7]. However, the passivity controller does not provide good tracking performance. To improve its tracking performance, a number of modifications to the passivity controller have also been proposed in the literature [8]. Other than passivity, adaptive control theory has also been used to design bilateral controllers [9]. Sliding mode [10] and H-infinity control [11] theories have also played an important role in robustifying the bilateral controllers. The use of intelligent techniques such as fuzzy logic and neural networks has also been explored in designing bilateral controllers [12]-[14]. The approximation property of these systems is central to the design of bilateral controllers. Another way to use fuzzy logic is model based fuzzy approach which has proved to be an effective tool in the control of nonlinear systems [15]-[22] and its application to bilateral teleoperation systems has also been reported [23]-[25].

State convergence belongs to the class of non-passive schemes and provides a simple and elegant method to design the bilateral controller [26]. The distinct feature of the method is the possibility of achieving desired dynamic behavior of the teleoperation system. This method has been shown to control both linear and nonlinear teleoperation systems in the absence and presence of time delays [27]. While good position tracking and the desired dynamic behavior was achieved with the originally proposed state convergence method, human perception of the remote environment was not good. To overcome this limitation, a modification to the original state convergence scheme has recently been proposed which has helped in improving the transparency of the teleoperation system while maintaining stability and the desired dynamic behavior [28]. The modified scheme

is named as transparency optimized state convergence method and will be used in this study.

This paper proposes to employ transparency optimized state convergence method in controlling a nonlinear teleoperation system which can be approximated by a class of TS fuzzy systems having common input and output matrices. A suitable form of PDC type fuzzy control law is selected to close the feedback loops around master and slave systems. The beauty of the selected control law lies in its capability to fully utilize the method of state convergence while providing large range operation. In our earlier work, we have proved that this control law can successfully establish the state convergence behavior in a nonlinear teleoperation system [24]-[25]. Following the same lines, we show that the fuzzy transparent bilateral controller can indeed be designed for the transparency optimized state convergence architecture with a nonlinear plant model. The validity of the proposed controller is confirmed through MATLAB simulations on a one DoF nonlinear teleoperation system.

The organization of the paper is as follows: We start by reviewing the transparency optimized state convergence architecture in Section 2. The proposed fuzzy logic controller is detailed in Section 3. Results are presented in Section 4. Section 5 draws the conclusion and provides future directions.

2 Transparency Optimized State Convergence Scheme

Transparency optimized state convergence scheme is recently proposed for the bilateral teleoperation systems with time delay in the communication channel. It is a modified form of the original state convergence scheme where the objectives of reflecting the full environmental force to the operator and the desired dynamic behavior of the closed loop teleoperation could not be achieved at the same time. This restriction is resolved to some extent in the modified version on the expense of limiting the allowable time delay in the communication channel and constraining the achievable closed loop behavior. Similar to the standard state convergence scheme, transparency optimized state convergence scheme also considers the master and slave systems modeled on state space as:

$$\begin{aligned} \overset{\text{g}}{x}_z &= A_z x_z + B_z u_z \\ y_z &= C_z x_z \end{aligned} \tag{1}$$

Where the subscript z denotes either master (z=m) or slave (z=s) systems and various matrix entries in (1) are given as:

$$A_z = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ & & ... & & \\ 0 & 0 & 0 & ... & 1 \\ -a_{z1} & -a_{z2} & -a_{z3} & ... & -a_{zn} \end{pmatrix} \tag{2}$$

$$\begin{aligned} B_z &= \begin{pmatrix} 0 & 0 & ... & 0 & b_{z1} \end{pmatrix}^T \\ C_z &= \begin{pmatrix} 1 & 0 & ... & 0 & 0 \end{pmatrix} \end{aligned} \tag{3}$$

The block diagram of the transparency optimized state convergence scheme is shown in Figure 1 and various parameters forming the architecture are described below:

T: This scalar parameter represents the time delay offered by the communication channel.

F_m: This scalar parameter represents the force applied by the human operator onto the master system.

G_2: This scalar parameter measures the influence of the operator's force into the slave system.

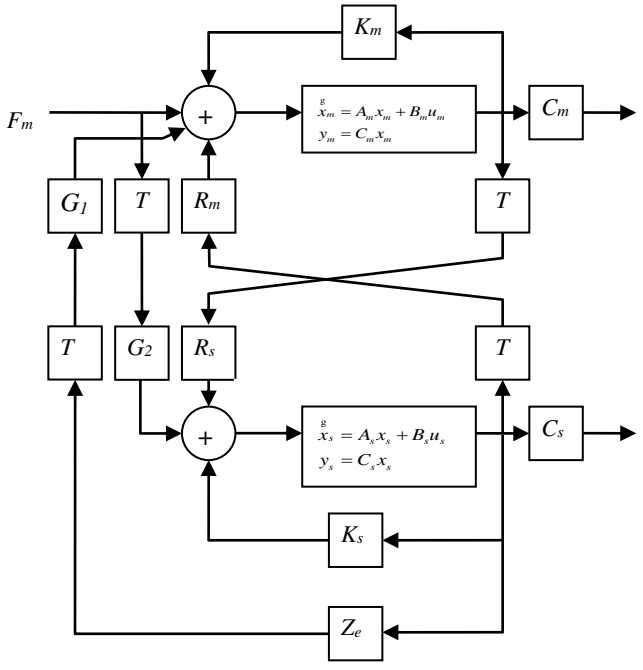


Figure 1
Transparent state convergence method

Z_e = [z_{e1},z_{e2},…,z_{en}]: This vector parameter is the model of the remote slave environment. When the environment is modeled by a spring-damper system, then this vector will contain two non-zero elements and rest of the elements will be zero.

G_1: This scalar parameter represents the influence of the environmental force when reflected onto the master system.

R_s = [r_{s1},r_{s2},…,r_{sn}]: This vector parameter represents the influence of the master's motion signals in the slave system.

R_m = [r_{m1},r_{m2},…,r_{mn}]: This vector parameter represents the influence of the slave's motion signals in the master system.

K_m = [k_{m1},k_{m2},…,k_{mn}]: This vector parameter is the state feedback controller for the master system.

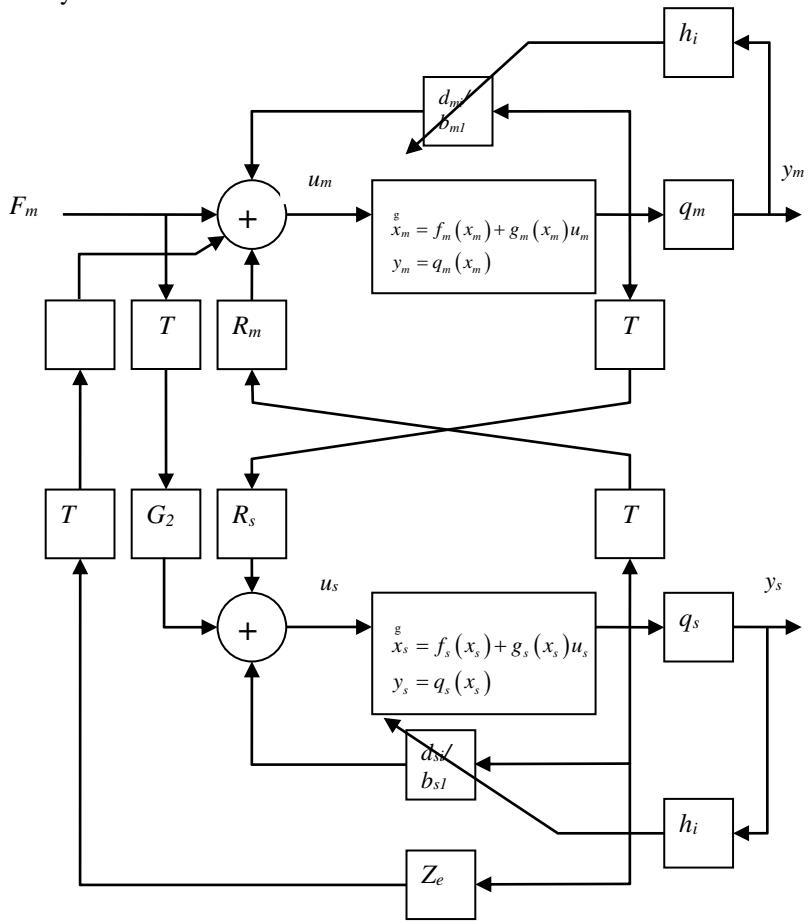


Figure 2
Prposed fuzzy transparent state convergence method

$K_s = [k_{s1}, k_{s2}, \ldots, k_{sn}]$: This vector parameter is the state feedback controller for the slave system.

Of these, G_1, G_2, R_s, R_m, K_m and K_s form $4n+2$ unknown parameters. The parameter G_1 can be freely chosen and is taken as unity when perfect transparency of the teleoperation system is desirable.

3 Proposed Transparent TS Fuzzy Logic Controller

In this section, we will show the development of a transparent TS fuzzy logic controller for a class of nonlinear teleoperation systems which can be approximated by TS fuzzy models in phase variable form with common input and output matrices. Such a nonlinear teleoperation system can be given as:

$$\begin{aligned} \overset{g}{x}_z &= f_z(x_z) + g_z u_z \\ y_z &= q_z x_z \end{aligned} \tag{4}$$

The TS fuzzy description of (1) with 'r' plant rules can be given as:

$$\begin{aligned} &\overset{g}{x}_{z1} = x_{z2} \\ &\overset{g}{x}_{z2} = x_{z3} \\ &\ldots \\ &\overset{g}{x}_{zn} = -\sum_{i=1}^{r} h_i(x_z) \sum_{j=1}^{n} a_{zij} x_{zj} + b_{z1} u_z \\ &y_z = x_{z1} \end{aligned} \tag{5}$$

Where $h_i(x_z)$ is the normalized firing strength of i^{th} fuzzy plant rule which is defined in (6) and satifies the two properties as in (6):

$$h_i(x_z) = \frac{\mu_i(x_z)}{\sum_{i=1}^{r} \mu_i(x_z)}, h_i(x_z) \geq 0, \sum_{i=1}^{r} h_i(x_z) = 1 \tag{6}$$

From Figure 2, we can introduce the TS fuzzy control law for the master system as:

$$u_m = \frac{1}{b_{m1}} \sum_{i=1}^{r} h_i(x_m) \sum_{j=1}^{n} d_{mij} x_{mj} + \sum_{j=1}^{n} \left(g_1 z_{ej} + r_{mj} \right) x_{sj}(t-T) + F_m \tag{7}$$

By plugging (7) in (5), the closed loop master sysem dynamics can be obtained as:

$$\overset{g}{x}_{mn} = \sum_{i=1}^{r} h_i(x_m) \sum_{j=1}^{n} \left(d_{mij} - a_{mij} \right) x_{mj} + b_{m1} \sum_{j=1}^{n} \left(g_1 z_{ej} + r_{mj} \right) x_{sj}(t-T) + b_{m1} F_m \tag{8}$$

Please note that we will considr n^{th} component of the system dynamics throughout the rest of the paper as in (8). Let us now introduce the time invariant coefficients for the master system as:

$$c_{mj} = d_{mij} - a_{mij} \tag{9}$$

With the coefficients in (9), the closed loop master system in (8) can be simplified as:

$$\overset{g}{x}_{mn} = \sum_{j=1}^{n} c_{mj} x_{mj} + b_{m1} \sum_{j=1}^{n} \left(g_1 z_{ej} + r_{mj} \right) x_{sj} \left(t - T \right) + b_{m1} F_m \tag{10}$$

Now, to close the loop around slave system, the following TS fuzzy control law is introduced (see Figure 2):

$$u_s = \sum_{i=1}^{r} h_i \left(x_s \right) \sum_{j=1}^{n} \left(\frac{d_{sij}}{b_{s1}} + z_{ej} \right) x_{sj} + \sum_{j=1}^{n} r_{sj} x_{mj} \left(t - T \right) + g_2 F_m \left(t - T \right) \tag{11}$$

The closed loop slave system can now be computed using (5) and (11) as:

$$\overset{g}{x}_{sn} = \sum_{i=1}^{r} h_i \left(x_s \right) \sum_{j=1}^{n} \left(d_{sij} - a_{sij} + b_{s1} z_{ej} \right) x_{sj} + b_{s1} \sum_{j=1}^{n} r_{sj} x_{mj} \left(t - T \right) + b_{s1} g_2 F_m \left(t - T \right) \tag{12}$$

Similar to the master system, we define the time invariant coefficients for the slave system as:

$$c_{sj} = d_{sij} - a_{sij} \tag{13}$$

With the definition in (13), the closed loop slave system in (12) can be written as:

$$\overset{g}{x}_{sn} = \sum_{j=1}^{n} \left(c_{sj} + b_{s1} z_{ej} \right) x_{sj} + b_{s1} \sum_{j=1}^{n} r_{sj} x_{mj} \left(t - T \right) + b_{s1} g_2 F_m \left(t - T \right) \tag{14}$$

As a part of the design procedure, state convergence method assumes the time delay in the communication channel to be small and the operator's force as constant. Thus, Taylor expansion of first order can be used to approximate the time delayed terms as:

$$\begin{aligned} &x_{mj} \left(t - T \right) ; \; x_{mj} - T \overset{g}{x}_{mj} \\ &x_{sj} \left(t - T \right) ; \; x_{sj} - T \overset{g}{x}_{sj} \\ &F_m \left(t - T \right) ; \; F_m - T \overset{g}{F}_m = F_m \end{aligned} \tag{15}$$

With the approximation in (15), closed loop master dynamics of (10) can be written as:

$$\overset{g}{x}_{mn} = \sum_{j=1}^{n} c_{mj} x_{mj} + b_{m1} \sum_{j=1}^{n} \left(g_1 z_{ej} + r_{mj} \right) x_{sj} - b_{m1} T \sum_{j=1}^{n-1} \left(g_1 z_{ej} + r_{mj} \right) \overset{g}{x}_{sj} - b_{m1} T \left(g_1 z_{en} + r_{mn} \right) \overset{g}{x}_{sn} + b_{m1} F_m \tag{16}$$

Similarly, closed loop dynamics of the slave system in (14) can be approximated as:

$$\overset{g}{x}_{sn} = \sum_{j=1}^{n} \left(c_{sj} + b_{s1} z_{ej} \right) x_{sj} + b_{s1} \sum_{j=1}^{n} r_{sj} x_{mj} - b_{s1} T \sum_{j=1}^{n-1} r_{sj} \overset{g}{x}_{mj} - b_{s1} T r_{sn} \overset{g}{x}_{mn} + b_{s1} g_2 F_m \tag{17}$$

By plugging (17) in (16) and using the phase variable representation of the slave system, closed loop master system dynamics of (16) can be written as:

$$\overset{g}{x}_{mn} = \frac{1}{\left(1 - T^2 b_{m1} b_{s1} r_{sn} \left(g_1 z_{en} + r_{mn} \right) \right)} \begin{pmatrix} \sum_{j=1}^{n} \left(c_{mj} - T b_{m1} b_{s1} \left(g_1 z_{en} + r_{mn} \right) r_{sj} \right) x_{mj} + \\ \sum_{j=1}^{n} \left(b_{m1} \left(g_1 z_{ej} + r_{mj} \right) - T b_{m1} \left(g_1 z_{en} + r_{mn} \right) \left(c_{sj} + b_{s1} z_{ej} \right) \right) x_{sj} - \\ T b_{m1} \sum_{j=1}^{n-1} \left(g_1 z_{ej} + r_{mj} \right) x_{sj+1} + T^2 b_{m1} b_{s1} \left(g_1 z_{en} + r_{mn} \right) \sum_{j=1}^{n-1} r_{sj} x_{mj+1} + \\ \left(b_{m1} - T b_{m1} b_{s1} g_2 \left(g_1 z_{en} + r_{mn} \right) \right) F_m \end{pmatrix} \tag{18}$$

Similarly, by plugging (16) in (17) and using the phase variable representation of the master system, the closed loop slave system dynamics of (17) can be written as:

$$\overset{g}{x}_{sn} = \frac{1}{\left(1 - T^2 b_{m1} b_{s1} r_{sn} \left(g_1 z_{en} + r_{mn} \right) \right)} \begin{pmatrix} \sum_{j=1}^{n} \left(b_{s1} r_{sj} - T b_{s1} r_{sn} c_{mj} \right) x_{mj} + \\ \sum_{j=1}^{n} \left(c_{sj} + b_{s1} z_{ej} - T b_{s1} b_{m1} r_{sn} \left(g_1 z_{ej} + r_{mj} \right) \right) x_{sj} - \\ T b_{s1} \sum_{j=1}^{n-1} r_{sj} x_{mj+1} + T^2 b_{s1} b_{m1} r_{sn} \sum_{j=1}^{n-1} \left(g_1 z_{ej} + r_{mj} \right) x_{sj+1} - \\ \left(b_{s1} g_2 - T b_{s1} b_{m1} r_{sn} \right) F_m \end{pmatrix} \tag{19}$$

Let us now define the state convergence error between master and slave systems as:

$$x_{ej} = x_{mj} - x_{sj}, \forall j = 1, 2, ..., n \tag{20}$$

We now write the closed loop master system dynamics of (18) in terms of state convergence error as:

$$\overset{g}{x}_{mn} = \frac{1}{\left(1-T^2 b_{m1} b_{s1} r_{sn}\left(g_1 z_{en}+r_{mn}\right)\right)} \begin{pmatrix} \sum_{j=1}^{n}\begin{pmatrix} c_{mj}-Tb_{m1}b_{s1}\left(g_1 z_{en}+r_{mn}\right)r_{sj}+b_{m1}\left(g_1 z_{ej}+r_{mj}\right)- \\ Tb_{m1}\left(g_1 z_{en}+r_{mn}\right)\left(c_{sj}+b_{s1}z_{ej}\right) \end{pmatrix} x_{mj} - \\ \sum_{j=1}^{n}\left(b_{m1}\left(g_1 z_{ej}+r_{mj}\right)-Tb_{m1}\left(g_1 z_{en}+r_{mn}\right)\left(c_{sj}+b_{s1}z_{ej}\right)\right)x_{ej} + \\ \sum_{j=1}^{n-1}\left(T^2 b_{m1}b_{s1}\left(g_1 z_{en}+r_{mn}\right)r_{sj}-Tb_{m1}\left(g_1 z_{ej}+r_{mj}\right)\right)x_{mj+1} + \\ Tb_{m1}\sum_{j=1}^{n-1}\left(g_1 z_{ej}+r_{mj}\right)x_{ej+1}+\left(b_{m1}-Tb_{m1}b_{s1}g_2\left(g_1 z_{en}+r_{mn}\right)\right)F_m \end{pmatrix} \tag{21}$$

Similarly, the closed loop slave system dynamics of (19) is also written in terms of state convergence error as:

$$\overset{g}{x}_{sn} = \frac{1}{\left(1-T^2 b_{m1} b_{s1} r_{sn}\left(g_1 z_{en}+r_{mn}\right)\right)} \begin{pmatrix} \sum_{j=1}^{n}\begin{pmatrix} c_{sj}+b_{s1}z_{ej}-Tb_{s1}b_{m1}r_{sn}\left(g_1 z_{ej}+r_{mj}\right)+ \\ b_{s1}r_{sj}-Tb_{s1}r_{sn}c_{mj} \end{pmatrix} x_{mj} + \\ \sum_{j=1}^{n}\left(c_{sj}+b_{s1}z_{ej}-Tb_{s1}b_{m1}r_{sn}\left(g_1 z_{ej}+r_{mj}\right)\right)x_{ej} + \\ \sum_{j=1}^{n-1}\left(T^2 b_{s1}b_{m1}r_{sn}\left(g_1 z_{ej}+r_{mj}\right)-Tb_{s1}r_{sj}\right)x_{mj+1} - \\ T^2 b_{s1}b_{m1}r_{sn}\sum_{j=1}^{n-1}\left(g_1 z_{ej}+r_{mj}\right)x_{ej+1}-\left(b_{s1}g_2-Tb_{s1}b_{m1}r_{sn}\right)F_m \end{pmatrix} \tag{22}$$

By taking the time derivative of (20) and using (21)-(22), we find the closed loop error dynamics of the teleoperation system as:

$$\overset{g}{x}_{en} = \frac{1}{\left(1-T^2 b_{m1} b_{s1} r_{sn}\left(g_1 z_{en}+r_{mn}\right)\right)} \begin{pmatrix} \sum_{j=1}^{n}\begin{pmatrix} c_{mj}-Tb_{m1}b_{s1}\left(g_1 z_{en}+r_{mn}\right)r_{sj}+b_{m1}\left(g_1 z_{ej}+r_{mj}\right)- \\ Tb_{m1}\left(g_1 z_{en}+r_{mn}\right)\left(c_{sj}+b_{s1}z_{ej}\right)-c_{sj}-b_{s1}z_{ej}+ \\ Tb_{s1}b_{m1}r_{sn}\left(g_1 z_{ej}+r_{mj}\right)-b_{s1}r_{sj}+Tb_{s1}r_{sn}c_{mj} \end{pmatrix} x_{mj} - \\ \sum_{j=1}^{n}\begin{pmatrix} c_{sj}+b_{s1}z_{ej}-Tb_{s1}b_{m1}r_{sn}\left(g_1 z_{ej}+r_{mj}\right)- \\ b_{m1}\left(g_1 z_{ej}+r_{mj}\right)+Tb_{m1}\left(g_1 z_{en}+r_{mn}\right)\left(c_{sj}+b_{s1}z_{ej}\right) \end{pmatrix} x_{ej} + \\ \sum_{j=1}^{n-1}\begin{pmatrix} T^2 b_{m1}b_{s1}\left(g_1 z_{en}+r_{mn}\right)r_{sj}-Tb_{m1}\left(g_1 z_{ej}+r_{mj}\right)- \\ T^2 b_{s1}b_{m1}r_{sn}\left(g_1 z_{ej}+r_{mj}\right)+Tb_{s1}r_{sj} \end{pmatrix} x_{mj+1} + \\ \sum_{j=1}^{n-1}\left(Tb_{m1}\left(g_1 z_{ej}+r_{mj}\right)+T^2 b_{s1}b_{m1}r_{sn}\left(g_1 z_{ej}+r_{mj}\right)\right)x_{ej+1} + \\ \left(b_{m1}-Tb_{m1}b_{s1}g_2\left(g_1 z_{en}+r_{mn}\right)-b_{s1}g_2+Tb_{s1}b_{m1}r_{sn}\right)F_m \end{pmatrix} \tag{23}$$

We now form an augmented system of the closed loop master-error systems' dynamics using (21) and (23) as:

$$\begin{pmatrix} \overset{\cdot}{x}_{mn} \\ \overset{\cdot}{x}_{en} \end{pmatrix} = \frac{1}{D}\sum_{j=1}^{n}\begin{pmatrix} (a_{11})_j & (a_{12})_j \\ (a_{21})_j & (a_{22})_j \end{pmatrix}\begin{pmatrix} x_{mj} \\ x_{ej} \end{pmatrix} + \frac{1}{D}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} F_m \tag{24}$$

Where the entry $(a_{xy})_j$ implies evaluation at j^{th} state and all the entries in (24) are given as (with the zeroth index values being zero: $r_{m0}=0, r_{s0}=0, z_{e0}=0$):

$$\begin{aligned}
a_{11} &= c_{mj} - Tb_{m1}b_{s1}\left(g_1 z_{en} + r_{mn}\right) r_{sj} + b_{m1}\left(g_1 z_{ej} + r_{mj}\right) - Tb_{m1}\left(g_1 z_{en} + r_{mn}\right)\left(c_{sj} + b_{s1} z_{ej}\right) + \\
&\quad T^2 b_{m1}b_{s1}\left(g_1 z_{en} + r_{mn}\right) r_{sj-1} - Tb_{m1}\left(g_1 z_{ej-1} + r_{mj-1}\right) \\
a_{12} &= b_{m1}\left(g_1 z_{ej} + r_{mj}\right) - Tb_{m1}\left(g_1 z_{en} + r_{mn}\right)\left(c_{sj} + b_{s1} z_{ej}\right) + Tb_{m1}\left(g_1 z_{ej-1} + r_{mj-1}\right) \\
a_{21} &= c_{mj} - Tb_{m1}b_{s1}\left(g_1 z_{en} + r_{mn}\right) r_{sj} + b_{m1}\left(g_1 z_{ej} + r_{mj}\right) - Tb_{m1}\left(g_1 z_{en} + r_{mn}\right)\left(c_{sj} + b_{s1} z_{ej}\right) - \\
&\quad c_{sj} - b_{s1} z_{ej} + Tb_{s1}b_{m1} r_{sn}\left(g_1 z_{ej} + r_{mj}\right) - b_{s1} r_{sj} + Tb_{s1} r_{sn} c_{mj} + T^2 b_{m1}b_{s1}\left(g_1 z_{en} + r_{mn}\right) r_{sj-1} - \\
&\quad Tb_{m1}\left(g_1 z_{ej-1} + r_{mj-1}\right) - T^2 b_{s1}b_{m1} r_{sn}\left(g_1 z_{ej-1} + r_{mj-1}\right) + Tb_{s1} r_{sj-1} \\
a_{22} &= c_{sj} + b_{s1} z_{ej} - Tb_{s1}b_{m1} r_{sn}\left(g_1 z_{ej} + r_{mj}\right) - b_{m1}\left(g_1 z_{ej} + r_{mj}\right) + Tb_{m1}\left(g_1 z_{en} + r_{mn}\right)\left(c_{sj} + b_{s1} z_{ej}\right) + \\
&\quad Tb_{m1}\left(g_1 z_{ej-1} + r_{mj-1}\right) + T^2 b_{s1}b_{m1} r_{sn}\left(g_1 z_{ej-1} + r_{mj-1}\right)
\end{aligned} \tag{25}$$

$$\begin{aligned}
b_1 &= b_{m1} - Tb_{m1}b_{s1} g_2\left(g_1 z_{en} + r_{mn}\right) \\
b_2 &= b_{m1} - Tb_{m1}b_{s1} g_2\left(g_1 z_{en} + r_{mn}\right) - b_{s1} g_2 + Tb_{s1}b_{m1} r_{sn} \\
D &= 1 - T^2 b_{m1}b_{s1} r_{sn}\left(g_1 z_{en} + r_{mn}\right)
\end{aligned} \tag{26}$$

According to the method of state convergence, error should evolve as an autonomous system. This will happen upon the satisfaction of the following conditions:

$$(a_{21})_j = 0, \forall j = 1, 2, ..., n \tag{27}$$

$$b_2 = 0 \tag{28}$$

Once the error will behave like an autonomous system, augmented system of (24) can be assigned the desired dynamic behavior. This leads to the following conditions:

$$\left(s - (a_{11})_j\right) \times \left(s - (a_{22})_j\right) = \left(s + p_j\right) \times \left(s + q_j\right), \forall j = 1, 2, ..., n \tag{29}$$

Where the coefficients p_j and q_j form the desired polynomials for master and error systems respectively:

$$\begin{aligned} &s^n + p_n s^{n-1} + ... + p_2 s + p_1 = 0 \\ &s^n + q_n s^{n-1} + ... + q_2 s + q_1 = 0 \end{aligned} \tag{30}$$

The design conditions (27)-(29) ensure that the states' error converges to zero and the master system exhibits the desired behavior. However, the convergence of force error is not guaranteed. To achieve that the operator force matches with the environmental force in steady state, we first compute the transfer function of the closed loop augmented system of (24) under the effect of autonomous error system:

$$\begin{aligned} &\frac{x_{mj}(s)}{F_m(s)} = \frac{(num)_j}{(den)_j}, \forall j = 1,2,...,n \\ &(num)_j = \begin{cases} (-1)^{n+1} s^{j-1} b_1, n > 2, \forall j = 1,2,...,n \\ s^{j-1} b_1, n = 2 \end{cases} \\ &(den)_j = s^n - (a_{11})_n s^{n-1} - (a_{11})_{n-1} s^{n-2} - ... - (a_{11})_2 s - (a_{11})_1 \end{aligned} \tag{31}$$

The transfer function in (31) can now be evaluated at steady state and compared against the stiffness of the environment as:

$$\frac{x_{m1}(0)}{F_m} = \frac{b_1}{-(a_{11})_1} = -\frac{1}{z_{e1}} \tag{32}$$

The design condition (32) ensures that the force error will converge to zero in steady state. Now, we have $4n+2$ design variables: $g_1, g_2, c_{mj}, c_{sj}, r_{mj}, r_{sj}, \forall j = 1,2,...,n$ while the number of design equations (27)-(29), (32) are $3n+2$. To create a balance, we let: $r_{mj} = -c_{mj}, \forall j = 1,2,...,n$ which will reduce the number of design variables to $3n+2$. However, to achieve that the environmental force is fully reflected to the operator, g_1 has to be unity which will again create an imbalance between the number of design variables and the design equations. To overcome this, the constant coefficient of the desired master system polynomial is constrained by the other teleoperation system's variables and now the design procedure is balanced.

4 Simulation Results

In order to validate the proposed fuzzy model based transparent controller, MATLAB simulations are carried out using a one DoF nonlinear teleoperation system which can be described in differential equation/state space form as:

$$J_z \ddot{\theta}_z^{g} + b_z \dot{\theta}_z^{g} + m_z g l_z \sin\theta_z = u_z$$
$$\begin{pmatrix} \dot{x}_{z1}^{g} \\ \dot{x}_{z2}^{g} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -a_{z1}\xi_z & -a_{z2} \end{pmatrix} \begin{pmatrix} x_{z1} \\ x_{z2} \end{pmatrix} + \begin{pmatrix} 0 \\ b_{z1} \end{pmatrix} u_z \tag{33}$$

Where x_{z1} and x_{z2} are the state variables representing the position and velocity of the master/slave systems, $a_{z1} = \frac{m_z g l_z}{J_z}$, $a_{z2} = \frac{b_z}{J_z}$, $b_{z1} = \frac{1}{J_z}$ and $\xi_z(t) = \frac{\sin x_{z1}(t)}{x_{z1}(t)}$ is the corresponding scheduling variable. To construct the TS fuzzy model of the teleoperation system, we determine the extreme values of the scheduling variable over the range of its operation, which is assumed to be $[-\pi/3 \quad \pi/3]$ in this study. The extreme values are found to be $\xi_{\min} = 0.827$ and $\xi_{\max} = 1.0$ which further help in constructing the following fuzzy sets:

$$\rho_1(\xi_z) = \begin{cases} 1 & , x_{z1} = 0 \\ \dfrac{\xi_z - \xi_{z\min}}{\xi_{\max} - \xi_{\min}} & , x_{z1} \neq 0 \end{cases}$$
$$\rho_2(\xi_z) = 1 - \rho_1(\xi_z) \tag{34}$$

Based on (34), the two rule TS fuzzy model of (33) can now be given as:

Model Rule 1: IF ξ_z is ρ_1 THEN $\begin{pmatrix} \dot{x}_{z1}^{g} \\ \dot{x}_{z2}^{g} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -a_{z11} & -a_{z12} \end{pmatrix} \begin{pmatrix} x_{z1} \\ x_{z2} \end{pmatrix} + \begin{pmatrix} 0 \\ b_{z11} \end{pmatrix} u_z$ (35)

Model Rule 2: IF ξ_z is ρ_2 THEN $\begin{pmatrix} \dot{x}_{z1}^{g} \\ \dot{x}_{z2}^{g} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -a_{z21} & -a_{z22} \end{pmatrix} \begin{pmatrix} x_{z1} \\ x_{z2} \end{pmatrix} + \begin{pmatrix} 0 \\ b_{z21} \end{pmatrix} u_z$ (36)

By assuming: m_m =0.5kg, l_m=0.5m, J_m=0.0417Kg-m^2, m_s =2.0kg, l_s=1.0m, J_s=0.667Kg-m^2; the parameters in (35) and (36) for master and slave systems are computed as:

$$a_{m11} = 58.86, a_{m12} = 12.0, b_{m11} = 24.0$$
$$a_{m21} = 48.67, a_{m22} = 12.0, b_{m21} = 24.0 \tag{37}$$

$$a_{s11} = 29.43, a_{s12} = 1.50, b_{s11} = 1.50$$
$$a_{s21} = 24.34, a_{s22} = 1.50, b_{s21} = 1.50 \tag{38}$$

Besides system parameters, we need environmental model and the desired polynomials for master and error systems to obtain the control gains. The

environment is assumed to behave like a spring-damper system with the following parameters:

$$Z_e = -\begin{pmatrix} 0.8 & 0.02 \end{pmatrix} \tag{39}$$

Also, the desired polynomials for master and error systems are selected as:

$$\begin{aligned} p(s) &: s^2 + 10s + p_1 = 0 \\ q(s) &: s^2 + 10s + 25 = 0 \end{aligned} \tag{40}$$

Please note that the last coefficient p_1 of the desired master polynomial cannot be chosen freely as it is constrained by other parameters and will be determined as a part of the solution. It is pertinent to mention that the selection of the desired polynomials in (40) is vital to ensure the stability and performance of the time delayed closed loop teleoperation system. Now, by considering the time delay in the communication channel to be $T = 0.01s$ and g_1 as unity, we obtain the following solution to the design equations (27)-(29),(32) through MATLAB symbolic toolbox:

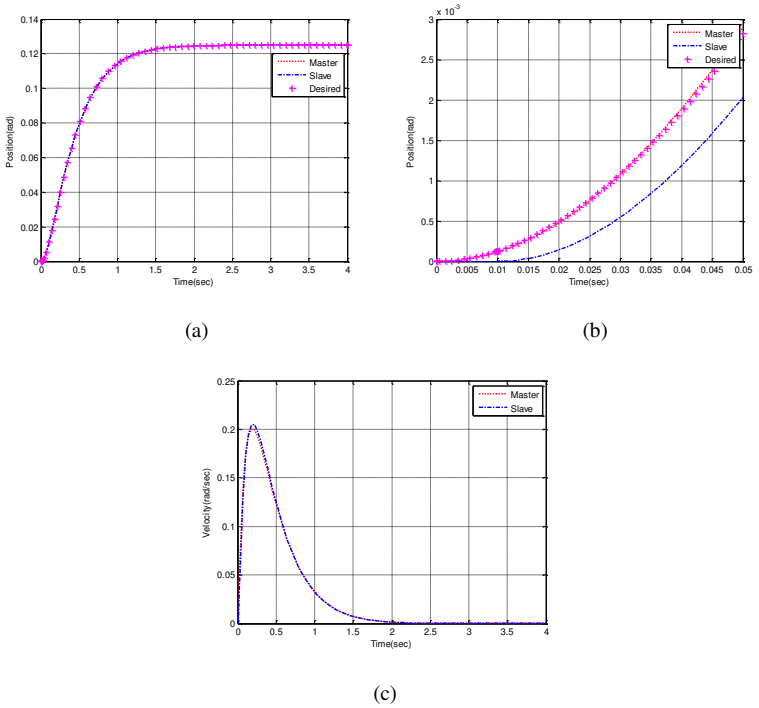
$$\begin{aligned} p_1 &= 21.2416 \\ g_2 &= 19.8151 \\ C_m &= \begin{pmatrix} c_{m1} & c_{m2} \end{pmatrix} = \begin{pmatrix} 0 & 0.3805 \end{pmatrix} \\ C_s &= \begin{pmatrix} c_{s1} & c_{s2} \end{pmatrix} = \begin{pmatrix} -50.5539 & -22.8256 \end{pmatrix} \\ R_m &= \begin{pmatrix} r_{m1} & r_{m2} \end{pmatrix} = \begin{pmatrix} 0 & -0.3805 \end{pmatrix} \\ R_s &= \begin{pmatrix} r_{s1} & r_{s2} \end{pmatrix} = \begin{pmatrix} 18.6505 & 7.9608 \end{pmatrix} \end{aligned} \tag{41}$$

In order to implement the fuzzy logic controllers on master and slave systems, we determine the control gains based on the solution in (41), system parameters in (37)-(38) and the time invariant parameters in (9),(13) as:

$$\begin{aligned} D_{m1} &= \begin{pmatrix} d_{m11} & d_{m12} \end{pmatrix} = \begin{pmatrix} 58.86 & 12.38 \end{pmatrix} \\ D_{m2} &= \begin{pmatrix} d_{m21} & d_{m22} \end{pmatrix} = \begin{pmatrix} 48.6768 & 12.38 \end{pmatrix} \\ D_{s1} &= \begin{pmatrix} d_{s11} & d_{s12} \end{pmatrix} = \begin{pmatrix} -21.1239 & -21.3256 \end{pmatrix} \\ D_{s2} &= \begin{pmatrix} d_{s21} & d_{s22} \end{pmatrix} = \begin{pmatrix} -26.2155 & -21.3256 \end{pmatrix} \end{aligned} \tag{42}$$

We now simulate the nonlinear teleoperation system of (33) by using the system parameters in Table 1 and the control gains in (41)-(42). The behavior of the teleoperation system as the operator applies a constant force of 0.1N is depicted in Figure 3. It can be observed that slave system is following the trajectory of the master system starting from the same initial conditions. Also, observe that the master system exhibits the desired dynamic behavior as assigned through the polynomial $p(s)$. The control inputs for both the master and slave systems in this

case are also recorded and displayed in Figure 4. It should be noted that the transparency optimized state convergence scheme is found to be sensitive to actuator saturation phenomenon and can easily be driven to instability. Thus, poles of the closed loop teleoperation system should be selected carefully to avoid the actuator saturation problem. We also analyze the environmental force which is reflected onto the master system. Figure 5 shows the operator's applied force as well as the force reflected from the slave system as it interacts with the environment. It can be observed that the operator is able to fully perceive the environment during the steady state. This result coincides with the design condition (32) which only ensures the force tracking in steady state. The behavior of the closed loop teleoperation is also analyzed under the application of more realistic time varying operator's force. The position and force tracking results for this case are shown in Figure 6. It can be observed that the slave system is at different initial position than the master system and is able to catch the master system after a transient. However, a constant force reflection error is observed during the ramp period, which disappears when the force becomes constant.



(a) (b)

(c)

Figure 3
Master and slave systems' states under an operator's force of 0.1N (a) Position signals (b) Zoomed position signals (c) Velocity signals

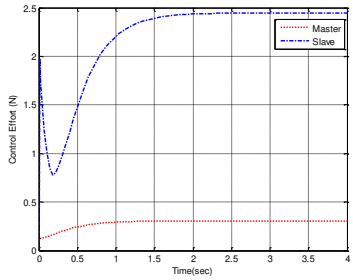


Figure 4
Control inputs for master and slave systems

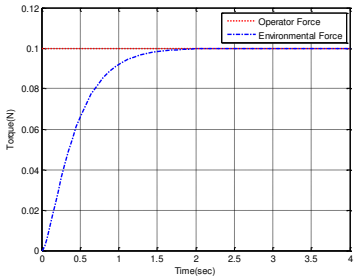
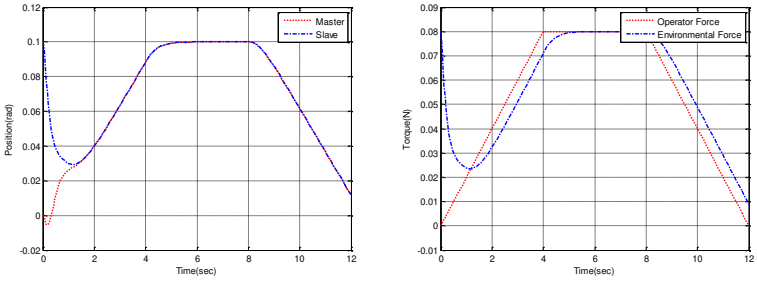


Figure 5
Operator and environmental forces

We also analyze the effect of uncertainties on the performance of the closed loop teleoperation system. To this end, we first consider 100% uncertainty in the time delay. With the same control gains as in (41)-(42) and considering $T = 0.02s$, we simulate the teleoperation system under the application of a constant applied force measuring 0.1N and the results are shown in Figure 7. It can be seen that although the state convergence between master and slave systems is achieved, a deviation from the desired dynamic behavior is evident. Further, this deviation increases as the uncertainty in the time delay increases which can be observed from figure 8 where a 200% uncertainty in the time delay ($T = 0.03s$) is considered. However, the transparency of the teleoperation system is achieved in steady state as both the slave position signal and the environmental force matches with the master position signal and the operator's applied force after the transient period.



(a) (b)

Figure 6
Performance of teleoperation system with time varying operator's force (a) Master-slave position signals (b) Operator-environment forces

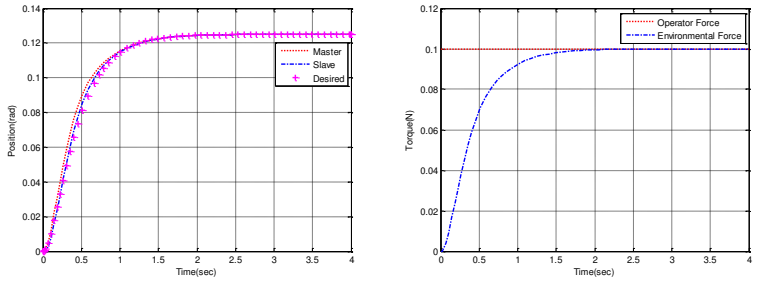
Besides the uncertainty in the time delay, we also study the effect of uncertainty in coefficient of viscous friction on the performance of teleoperation system. The result of this analysis is shown in Figure 9 where a 50% uncertainty is considered in the coefficient of viscous friction of both master and slave systems. It can be

seen that the teleoperation system is exhibiting a sluggish response. The effect of adding the uncertainty in the time delay and the coefficient of viscous friction at the same time is shown in Figure 10. It is evident that the uncertainty in the coefficient of viscous friction has a greater impact in degrading the system's performance.

Finally, we compare the performance of the proposed fuzzy logic controller with the existing linear controller in achieving the transparency during large range operation. The linear controller is derived from the proposed controller using (43) and is given in (44).

$$k_{mj} = \frac{1}{b_{m1}}\left(a_{mj} + c_{mj}\right), \forall j = 1,2,...,n$$
$$k_{sj} = \frac{1}{b_{s1}}\left(a_{sj} + c_{sj}\right), \forall j = 1,2,...,n \tag{43}$$

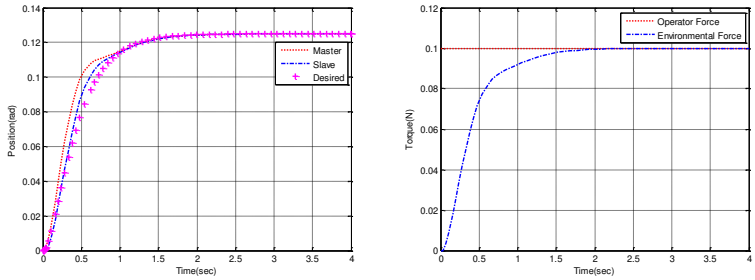
$$K_m = \begin{pmatrix} k_{m1} & k_{m2} \end{pmatrix} = \begin{pmatrix} 2.4525 & 0.5159 \end{pmatrix}$$
$$K_s = \begin{pmatrix} k_{s1} & k_{s2} \end{pmatrix} = \begin{pmatrix} -14.0826 & -14.2171 \end{pmatrix} \tag{44}$$



(a) (b)

Figure 7

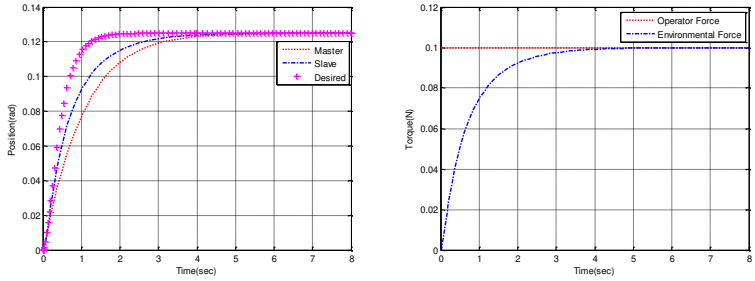
Effect of 100% uncertainty in time delay on the performance of teleoperation system (a) Master-slave position signals (b) Operator-environment forces



(a) (b)

Figure 8

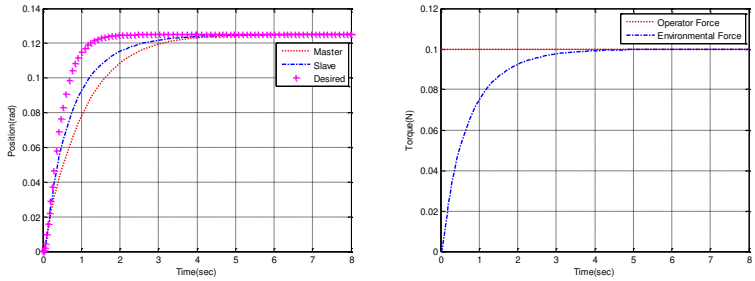
Effect of 200% uncertainty in time delay on the performance of teleoperation system (a) Master-slave position signals (b) Operator-environment forces



(a) (b)

Figure 9

Effect of 50% uncertainty in the coefficient of viscous friction on the performance of teleoperation system (a) Master-slave position signals (b) Operator-environment forces

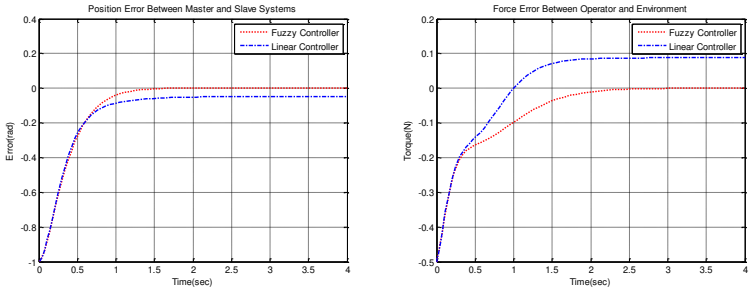


(a) (b)

Figure 10

Effect of 50% uncertainty in the coefficient of viscous friction and time delay on the performance of teleoperation system (a) Master-slave position signals (b) Operator-environment forces

By considering the final position to be reached as 1 rad, the nonlinear teleoperation system is simulated and the performance of the two controllers is recorded. The errors in the position and the force signals for the two cases are then computed and are displayed in Figure 11. It can be observed that the proposed fuzzy logic controller has shown superior performance as the error in position and force signals has converged to zero while a constant position and force error is seen in case of the linear controller. Thus, perfect transparency is achieved by the proposed fuzzy logic controller.



(a) (b)

Figure 11

Comparison of proposed fuzzy logic controller and existing linear controller (a) Master-slave position errors (b) Operator-environment force errors

Conclusions

This paper has presented the design of a fuzzy model based transparent controller for a nonlinear teleoperation system based on its TS fuzzy description. The proposed TS fuzzy logic control laws for the master and slave systems allow using the method of state convergence in its true sense. The feasibility of the presented approach is evaluated through simulations in MATLAB environment on a one DoF tele-manipulator. It is concluded that the proposed approach can control a nonlinear teleoperation system with a small time delay in the communication channel. Future work involves enabling the scheme to work in the presence of time varying delays. The robustness of the scheme to parameter uncertainties need to be improved as well.

References

[1] M. Ferre, M. Buss, R. Aracil, C. Melchiorri, C. Balaguer: Advances in Telerobotics, Springer, 2007

[2] P. F. Hokayem, M. W. Spong: Bilateral Teleoperation: an Historical Survey, Automatica, Vol. 42, 2006, pp. 2035-2057

[3] R. Muradore, P. Fiorini: A Review of Bilateral Teleoperation Algorithms, Acta Polytechnica Hungarica, Vol. 13, No. 1, 2016, pp. 191-208

[4] J. Artigas, G. Hirzinger: A Brief History of DLR's Space Telerobotics and Force Feedback Teleoperation, Acta Polytechnica Hungarica, Vol. 13, No. 1, 2016, pp. 239-249

[5] R. Anderson, M. W. Spong: Bilateral Control of Teleoperators with Time Delay: IEEE Transactions on Automatic Control, Vol. 34, No. 5, 1989, pp. 494-501

[6] G. Niemeyer, J. J. E. Slotine: Stable Adaptive Teleoperation, IEEE Journal of Oceanic Engineering, Vol. 16, No. 1, 1991, pp. 152-162

[7] J. Ryu, D. Kwon, B. Hannaford: Stable Teleoperation with Time-domain Passivity Control, IEEE Transactions on Robotics and Automation, Vol. 20, No. 2, 2004, pp. 365-373

[8] A. Ziminejad, M. Tavakoli, R. V. Patel, M. Moallem: Transparent Time-delayed Bilateral Teleoperation Using Wave Variables, IEEE Transactions on Control Systems Technology, Vol. 16, No. 3, 2008, pp. 548-555

[9] L. Chan, F. Naghdy, D. Stirling: Application of Aadaptive Controllers in Teleoperation Systems, IEEE Transactions on Human-Machine Systems, Vol. 44, No. 3, 2014, pp. 337-352

[10] A. Hace, M. Franc: FPGA Implementation of Sliding Mode Control Algorithm for Scaled Bilateral Teleoperation, IEEE Transactions on Industrial Electronics, Vol. 9, No. 3, 2013, pp. 1291-1300

[11] M. Boukhnifer, A. Ferreira: H∞ Loop Shaping Bilateral Controller for a Two-fingered Tele-micromanipulation System, IEEE Transations on Control Systems Technology, Vol. 15, No. 5, 2007, pp. 891-905

[12] V. T. Minh, F. M. Hashim: Adaptive Teleoperation System with Neural Network Based Multiple Model Control, Mathematical Problems in Engineering, Vol. 2010, pp. 1-16

[13] J. Zhu, W. Gueaieb: Adaptive Fuzzy Logic Control for Time-delayed Bilateral Teleoperation, Autonomous and Intelligent Systems, Vol. 7326, 2012, pp. 106-115

[14] Y. Yang, C. Hua, X. Guan: Adaptive Fuzzy Finite-time Coordination Control for Networked Nonlinear Bilateral Teleoperation System, IEEE Transactions on Fuzzy Systems, Vol. 22, No. 3, 2014, pp. 631-641

[15] H. O. Wang, K. Tanaka, M. F. Griffin: An Approach to Fuzzy Control of Nonlinear Systems: Stability and Design Issues, IEEE Transactions on Fuzzy Systems, Vol. 4, No. 1, pp.14-23, 1996

[16] L. K. Wong, F. H. F. Leung, P. K. S. Tam: Design of Fuzzy Logic Controllers for Takagi-Sugeno Fuzzy Model Based System with Guaranteed Performance, International Journal of Approximate Reasoning, Vol. 30, pp. 41-55, 2002

[17] R-E. Precup, M-L. Tomescu, S. Preitl: Fuzzy Logic Control System Stability Analysis Based on Lyapunov's Direct Method, International Journal of Computers, Communications and Control, Vol. 4, 2009, pp. 415-426

[18] Y. Wei, J. Qiu, H. R. Karimi: Reliable Output Feedback Control of Discrete Time Fuzzy Affine Systems with Actuator Faults, IEEE Transactions on Circuits and Systems, Vol. 64, No. 1, 2017, pp. 170-181

[19] Y. Wei, J. Qiu, P. Shi, M. Chadli: Fixed-Order Piecewise-Affine Output Feedback Controller for Fuzzy-Affine-Model-Based Nonlinear Systems with Time Varying Delay, IEEE Transactions on Circuits and Systems, Vol. 64, No. 4, 2017, pp. 945-958

[20] Z. Lendek, A. Sala, P. Garcia, R. Sanchis: Experimental Application of Takagi-Sugeno Observers and Controllers in a Nonlinear Electromechanical System, Journal of Control Engineering and Applied Informatics, Vol. 15, No. 4, 2013, pp. 3-14

[21] T-C. Lin, T-Y. Lee, V. E. Balas: Adaptive Fuzzy Sliding Mode Control for Synchronization of Uncertain Fractional Order Chaotic Systems, Chaos, Solitons & Fractals, Vol. 44, 2011, pp. 791-801

[22] U. Farooq, J. Gu, M. E. El-Hawary, V. E. Balas, M. U. Asad: Experimental Study of Optimal Takagi Sugeno Fuzzy Controller for Rotary Inverted Pendulum, Proc. IEEE FUZZ, 2015, pp. 1-7

[23] X. Yang, C. C. Hua, J. Yan, X. P. Guan: A New Master-Slave Torque Design for Teleoperation System by T-S Fuzzy Approach, IEEE Transactions on Control Systems Technology, Vol. 23, No. 4, 2015, pp. 1611-1619

[24] U. Farooq, J. Gu, M. El-Hawary, M. U. Asad, G. Abbas: Fuzzy Model Based Bilateral Control Design of Nonlinear Tele-operation System Using Method of State Convergence, IEEE Access, Vol. 4, 2016, pp. 4119-4135

[25] U. Farooq, J. Gu, M. El-Haway, M. U. Asad, Faiqa Rafiq: Transparent Fuzzy Bilateral Control of a Nonlinear Teleoperation System Through State Convergence, Proc. IEEE International Conference on Emerging Technologies, 2016, pp. 1-7.

[26] J. M. Azorin, O. Reinoso, R. Aracil, M. Ferre: Generalized Control Method by State Convergence of Teleoperation Systems with Time Delay, Automatica, Vol. 40, No. 9, 2004, pp. 1575-1582

[27] M. A. Artigas, R. Aracil, M. Ferre, C. Garcia: Bilateral Adaptive Control by State of Convergence in Teleoperation Systems, Proc. IEEE International Conference on Control Applications, 2009, pp. 548-553

[28] R. Aracil, J. M. Azorin, M. Ferre, C. Pena: Bilateral Control by State Convergence Based on Transparency for Systems with Time Delay, Robotics and Autonomous Systems, Vol. 61, No. 2, 2013, pp. 86-94

Internet-based Bilateral Teleoperation Using a Revised Time-Domain Passivity Controller

Lőrinc Márton[1], Zoltán Szántó[1], Tamás Haidegger[2], Péter Galambos[2], József Kövecses[3]

[1]Department of Electrical Engineering, Sapientia Hungarian University of Transylvania, Tirgu Mures, Romania, E-mail: martonl@ms.sapientia.ro

[2]Antal Bejczy Center for Intelligent Robotics, Óbuda University, Budapest, Hungary

[3]Department of Mechanical Engineering and the Centre for Intelligent Machines, McGill University, Montreal, Canada

Abstract: This study presents a teleoperation system for remote control of mobile manipulators over the Internet. A bilateral control algorithm is proposed that can assure both stability and proper force reflection in the presence of non-constant delay in the communication channels between the master and the slave. The control approach in this paper is based on the time domain passivity concept and proposes a modified passivity controller to assure enhanced transparency with bounded control actions in the presence of time-varying communication delay. Transatlantic and inter-European bilateral teleoperation experiments are also reported (Montreal, Canada - Tirgu Mures, Romania; Budapest, Hungary - Tirgu Mures, Romania). The experimental measurements show the applicability of the control approach and its benefits on the teleoperation performances.

Keywords: Telerobotics; Internet; Networked control systems; Delay systems; Passivity

1 Introduction

The Internet represents the common communication medium for the implementation of such teleoperation applications that have a significant physical distance between the human operator and the teleoperated robot. For such cases when the teleoperated robot is mobile, wireless links should also be included in the communication channel. These wireless hops in most cases represent the critical link regarding the communication performances. The communication channel's performance can severely be compromised, e.g. by other data transfer channels sharing the same wireless communication medium, or by wireless signal strength change, which appears in the case of mobile teleoperation.

The communication performance degradation manifests in increased delay (communication lag) and delay variation [24].

In teleoperation systems, the desired position and velocity of the remote robot is generated by the human operator using a haptic device (master). During bilateral teleoperation, the controlled robot (slave) sends back to the master haptic information about forces or torques that influence the robot's motion, allowing the human operator to feel the effect of the environment-robot contact during the teleoperation.

A major challenge during the control system design for bilateral teleoperation systems is to guarantee stability under any operating circumstances and, simultaneously, to assure a reliable position tracking and accurate force reflection (transparency) under various communication conditions [4], [7].

Several approaches were proposed to implement telerobotic systems over the Internet, such as the Plugfest experiment [10] (unilateral teleoperation) or teleoperation over the PlanetLab overlay network [1]. In the work [9], a control and communication co-design approach was proposed for the distant teleoperation of drive by wireless vehicles.

Such extended PD (Proportional - Derivative)-like control algorithms that can also assure the passivity of delayed bilateral teleoperation systems were introduced, e.g. in [16]. The dissipation parameters of such controllers are designed in function of the upper bound of the communication lag [15]. A similar concept was applied in the paper [8] considering discrete-time controllers. Generally, for greater delay values greater dissipation gain is required in the control algorithm. The Passive-Set-Position-Modulation approach [11] also applies a PD-like algorithm for the distantly teleoperated robot. This method modulates the position set-point such to assure as good position tracking performance as possible without violating the passivity of the teleoperation system.

The wave variable based bilateral control [3] is also applicable for such teleoperation systems in which communication delay is non-constant. With this approach, the stability can be assured if the variation of the delay is upper bounded.

In the case of the time domain passivity control the controller term that assures stability is activated only when an observer indicates that the passivity of the system is compromised [19]. The time domain passivity approach does not necessitate information about the communication lag, but it can be seen that the greater the delay the more energy is dissipated by the passivity controller both on the master and the slave side, as it was discussed in [14]. The paper [2] extends the concept of r-passivity to time domain passivity control schemes to assure precise position tracking and meanwhile to preserve the stability of the teleoperation in the presence of time-varying communication lag. The paper [13] combines the time domain passivity control approach with the identification of the

remote environment's parameters to achieve stable force tracking. The time domain passivity control approach is also applicable for teleoperated mobile robots [18].

Along with the clear advantage of the time domain passivity control that it modifies the force and velocity signals only when the passivity condition is violated, it also has two shortages. The first is related to the magnitude of the control signal: as was already mentioned in the early paper [6], for near-zero force and velocity values the control signals can take such large values that compromise the implementability of the passivity controller. Second, the control signal can show oscillatory behavior. This phenomenon was already reported in previous works (see, e.g. the experimental measurements presented in [18] or [19].) The switching is accentuated, when the communication delay is time-varying [14], as in the case of Internet-based teleoperation. To deal with these shortcomings, this paper proposes a modified time domain passivity control which assures enhanced force reflection in the presence of non-constant communication lag with bounded control actions. Our first attempt to formulate a time domain passivity controller with bounded control signals can be found in the conference paper [14]. Another approach to solve the unbounded control signal problem was proposed in the paper [25]. However, in this paper the authors assumed that the upper bound of the communication delay variation is known.

In this current work, we give a more general form of the revised time domain passivity controller and we also perform the analysis of it. New realistic bilateral teleoperation experimental results are also presented to support the theoretical results. Teleoperation experiments over the Internet using commercial mobile manipulator and haptic devices between robotics laboratories situated in different countries (Hungary-Romania and Canada-Romania respectively) are presented.

The rest of the paper is organized as follows: before introducing the proposed improved control algorithm, the time domain passivity controller is invoked in Section 2. Section 3 presents the new theoretical results, i.e. the proposed control algorithm and its analysis. Experimental measurements are presented in Section 4. Finally, Section 5 concludes this paper.

2 Reminder on Time-Domain Passivity Control Approach for Delayed Bilateral Teleoperation Systems

This section recalls the original time domain passivity control approach (see, e.g. also [5]) which will be developed further in the next section.

A typical bilateral teleoperation system is shown in Fig. 1. The user (H) generates position/velocity commands using the haptic master device (M). The notation v_m stands for the master-side velocity. This command is sent through the communication medium to the slave robot's control system. On the slave-side the delayed v_m value is applied as the reference for the robot controller (SC). The robot controller has to ensure that the slave robot (S) tracks the received master-side velocity value. When there is direct contact between slave robot and the environment (E), the slave-side force f_s shows the effect of the environment on the slave robot. In Fig. 1, v_s and f_m represent the slave-side velocity and master-side force. In the communication channels d_{ms} and d_{sm} denote the time-varying communication lags.

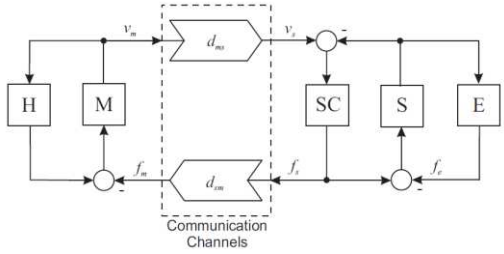


Figure 1
Block scheme of teleoperation system

The time domain passivity controller is designed such to compensate the effect of the communication delay on the stability of the closed loop teleoperation system. During the controller design, it is assumed that both the master side (master robot controlled by the human operator) and the slave side (slave robot that can be in contact with the environment) are passive. As the power-continuous interconnections of passive systems is also a passive system, if the passivity of the ``Communication Channels" can be guaranteed, the bilateral teleoperation system is passive. The *energetic passivity* condition for the communication medium has the form:

$$E(t) = \int_0^t \left(f_m(\tau) v_m(\tau) + f_s(\tau) v_s(\tau) \right) d\tau \geq 0, \forall t > 0, \tag{1}$$

where $E(t)$ is the energy measured at the ports of the ``Communication Channels" block. It is considered that $E(0)=0$.

According to the time domain passivity concept the passivity controller, which is responsible for the energy dissipation is active only when passivity condition (1) is not satisfied. The energy at the ports of the ``Communication Channels" (see Fig. 1) is computed by using the numerical approximation [20]:

$$E[k] = E_m[k] + E_s[k] = T\sum_{i=1}^{k}\left(f_m[i]v_m[i] + f_s[i]v_s[i]\right). \tag{2}$$

Here $T>0$ represents the sampling period used for the approximation. E_m stands for energy computed the master-side port, and E_s is the energy computed at the slave-side port.

When there is a communication lag between the master- and slave-side, the energy function (2) cannot be computed: at the master-side port there is no information about the slave-side velocity and force at the kth sample because of the delay, and vice versa.

This problem can be solved by decomposing of the computed energy into input (*IN*) and output (*OUT*) energies at both ports [19]:

$$E_s^{IN}[k] = \begin{cases} E_s^{IN}[k-1] + Tf_s[k]v_s[k], \text{if } f_s[k]v_s[k] > 0, \\ E_s^{IN}[k-1], \text{otherwise}, \end{cases} \tag{3}$$

$$E_s^{OUT}[k] = \begin{cases} E_s^{OUT}[k-1] - Tf_s[k]v_s[k], \text{if } f_s[k]v_s[k] < 0, \\ E_s^{OUT}[k-1], \text{otherwise}, \end{cases} \tag{4}$$

$$E_m^{IN}[k] = \begin{cases} E_m^{IN}[k-1] + Tf_m[k]v_m[k], \text{if } f_m[k]v_m[k] > 0, \\ E_m^{IN}[k-1], \text{otherwise}, \end{cases} \tag{5}$$

$$E_m^{OUT}[k] = \begin{cases} E_m^{OUT}[k-1] - Tf_m[k]v_m[k], \text{if } f_m[k]v_m[k] < 0, \\ E_m^{OUT}[k-1], \text{otherwise}, \end{cases} \tag{6}$$

As it was presented in [19], the ``Communication Channels" module is passive if

$$\Delta E_m[k] \le 0 \text{ and } \Delta E_s[k] \le 0 \tag{7}$$

$$\Delta E_m[k] = E_m^{OUT}[k] - E_m^{IN}[k - d_M[k]] \tag{8}$$

$$\Delta E_s[k] = E_s^{OUT}[k] - E_s^{IN}[k - d_s[k]] \tag{9}$$

where d_M and d_S are the discrete-time delays that are calculated as: $d_M = d_{sm}/T$, $d_S = d_{ms}/T$. The term (8) is computable at the master side and the expression (9) computable at the slave side.

When the passivity conditions (7) are not satisfied, the passivity controller adds to the signals received through the communication channels an extra energy-

dependent term. The role of this energy extra dissipation term is to assure the passivity of the "Communication Channels" module. On the master side, the Passivity Controller has the form [19]:

$$f_m[k] = f_s[k - d_M[k]] + \Delta f_m[k] \tag{10}$$

$$\Delta f_m[k] = \begin{cases} \dfrac{\Delta E_m[k]}{T v_m[k]}, \text{ if } v_m[k] \neq 0 \text{ and } \Delta E_m[k] > 0 \\ 0, \text{otherwise.} \end{cases} \tag{11}$$

$\Delta f_m[k]$ is an energy dissipation term which is active only when the condition (8) is not satisfied.

The passivity controller has a similar form at the slave side as well.

3 Enhanced Time-Domain Passivity Controller

3.1 The Proposed Control Law

The control signal given in (10) can take hardly implementable values when v_m takes near-zero values.

On the other hand, the force signal displayed to the human operator can oscillate in the presence of time-varying communication lag. This phenomenon appears because the second, damping term in (10) switches off for small delay values, but for large delay, it can take high values.

To handle these problems a novel passivity controller is introduced in this section. The master-side controller will be presented.

To avoid the excessive control values, replace the $1/v_m$ term in the control law (10) with a bounding function $\delta(v_m)$ which has the following properties:

- $\delta(v_m) \to 1/v_m$ if $|v_m|$ is large, and
- $\delta(v_m) \to 1/\delta_v \operatorname{sgn}(v_m)$ if $v_m \to 0$. Here $\delta_v > 0$.

An example for such a function is (see Fig. 2)

$$\delta(v_m) = \frac{1}{v_m + \delta_v \operatorname{sgn}(v_m)} \tag{12}$$

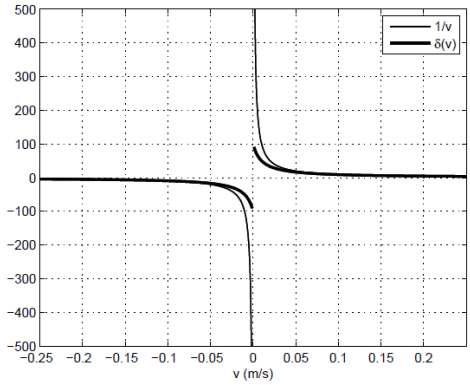


Figure 2
The $\delta(\cdot)$ function

The function $\delta(\cdot)$ assures the avoidance of the excessive control values, but in the low velocity regime the control law could dissipate less energy as the original control law. To tackle this problem, a modified energy term ($\Delta\hat{E}_m[k]$) is used in the control algorithm:

$$f_m[k] = f_s[k - d_M[k]] + \Delta f_m[k] \tag{13}$$

$$\Delta f_m[k] = \begin{cases} \delta(v_m[k])\dfrac{\Delta\hat{E}_m[k]}{T}, \text{if } \Delta E_m[k] > 0 \\ 0, \text{otherwise.} \end{cases} \tag{14}$$

The value of $\Delta\hat{E}_m[k]$ is formulated such that $\delta(v_m[k])\dfrac{\Delta\hat{E}_m[k]}{T} \to \dfrac{\Delta E_m[k]}{Tv_m[k]}$.

Accordingly, $\Delta E_m[k]$ can be generated by the following dynamic law:

$$\Delta\hat{E}_m[k+1] = \begin{cases} \Delta\hat{E}_m[k+1] + \dfrac{1}{T_I}\left(\Delta E_m[k] - v_m[k]\delta(v_m[k])\Delta\hat{E}_m[k]\right) \\ \quad \text{if } \Delta E_m[k] > v_m[k]\delta(v_m[k])\Delta\hat{E}_m[k] \\ \Delta E_m[k+1], \text{otherwise.} \end{cases} \tag{15}$$

Here $T_I > 0.5$ determines the convergence speed of dynamic equation above.

3.2 The Analysis of the Control Law

3.2.1 Steady-State Analysis

Proposition 1: If $v_m \neq 0$, the control law given by (13), (15) and (12) in steady-state is equivalent to the original control (10).

If $|v_m| > 0$ is constant, then the equation (15) is a first order discrete time filter with a pole $z = 1 - v_m \delta(v_m)/T_I$. As $0 \leq v_m \delta(v_m) < 1$ this filter is stable for $T_I > 0.5$.

If $\Delta\hat{E}_m[k+1] = \Delta\hat{E}_m[k]$, then $\Delta\hat{E}_m[k] = \dfrac{\Delta E_m[k]}{\delta(v_m[k])v_m[k]}$. With this value, the control (13) is equivalent to the control (10).

It can also be seen that the dynamics (15) has a low pass filter character, the filtered value of ΔE_m appears in the dissipation term of the control law; hence the high-frequency switching of the control signal is attenuated.

On the other hand, if $v_m = 0$, the dissipation term Δf_m in the original time domain passivity controller is unbounded. In the modified control (13), it increases with finite increments in each sampling period if $\Delta E_m = v_m \delta(v_m) \Delta\hat{E}_m$.

3.2.2 Passivity Analysis

To perform the analysis of time domain passivity control with bounded control signals the notion of *controller passivity* [12] is used. According to this condition, the maximum amount of energy generated by the network "Communication Channels" (see Fig. 1) that has the passivity controllers on its ports must be bounded. The controller passivity implies the energetic passivity of the bilateral teleoperation system if the master and slave side teleoperators are passive [12].

In the case of bilateral teleoperation systems that apply time domain passivity control the controller passivity condition is satisfied if:

- $\Delta E_m[k]$ is bounded for all k, and
- $\Delta E_m[k] > 0$ only for a finite number of sampling periods.

First, it is considered that, in concordance with the practical applications, the absolute values of the force, velocity and control signals in the bilateral teleoperation system are upper-bounded:

Assumption 1 $|v_m| \leq v_{MAX}$, $|v_s| \leq v_{MAX}$, $|f_m| \leq f_{MAX}$, $|f_s| \leq f_{MAX}$, $|\Delta f_m| \leq f_{MAX}$, $|\Delta f_s| \leq f_{MAX}$.

Second, the communication delays are assumed finite:

Assumption 2 $0 \leq d_{ms} < \infty$ and $0 \leq d_{sm} < \infty$.

Third, it is presumed that the controller passivity can be assured with bounded control signal.

Assumption 3 The bounded dissipation term ($|\Delta f_m| \leq f_{MAX}$) assures that the $\Delta E_m > 0$ only for a finite number of sampling periods.

Proposition 2 If the Assumptions 1, 2 and 3 hold, the Communication Network with the proposed controller (13) on its ports satisfies the controller passivity condition.

If $\Delta E_m[k] > v_m[k]\delta(v_m[k])\Delta\hat{E}_m[k]$, the energy term $\Delta\hat{E}_m[k]$ for all k is increasing, see the equation (15).

As $\Delta\hat{E}_m[k]$ is increasing and $\delta(v_m[k])$ is lower bounded for $|v_m| \leq v_{MAX}$, $|\Delta f_m|$ also increases, and it reaches $\min\{\Delta E_m/T|v_m|, f_{MAX}\}$ within a finite number of sampling periods. Hence, by Assumption 3, ΔE_m exceeds zero only for a finite number of sampling periods.

On the other hand, by Assumptions 1 and 2, ΔE_m is upper bounded in each sampling period.

3.3 Extension to Multi-DOF Teleoperators

The proposed algorithm can be extended to multi-DOF (Degree Of Freedom) master- and slave-side robots by applying the method presented in [17].

Consider that the master-side and the slave-side robots have the same DOF i.e., dim ($\mathbf{f}_m$)= dim ($\mathbf{v}_m$)= dim ($\mathbf{f}_s$)=dim ($\mathbf{v}_s$) =DOF. Then the energy at the ports of the Communication Network is computable as:

$$E[k] = E_m[k] + E_s[k] = T\sum_{i=1}^{k}\left(\mathbf{f}_m[i]^T\,\mathbf{v}_m[i] + \mathbf{f}_s[i]^T\,\mathbf{v}_s[i]\right) \tag{16}$$

Based on this relation, the master- and slave-side observers and the passivity condition can be formulated in the same way as in the equations (3)-(9).

The passivity controller (13) can be reformulated as:

$$\mathbf{f}_m[k] = \mathbf{f}_s[k - d_M[k]] + \begin{cases} \dfrac{\Delta\hat{E}_m[k]}{T(\delta_v + \mathbf{v}_m^T[k]\mathbf{v}_m[k]))} P_{\mathbf{f}_s}(\mathbf{v}_m[k]), \text{if } \Delta E_m[k] > 0 \\ 0, \text{otherwise.} \end{cases} \tag{17}$$

where $P_{\mathbf{f}_s}(\mathbf{v}_m[k])$ denotes the orthogonal projection of $\mathbf{v}_m$ onto $\mathbf{f}_s$ and δ_v>0. Here, the bounding function has the form

$$\delta(\mathbf{v}_m) = \frac{P_{\mathbf{f}_s}(\mathbf{v}_m[k])}{T(\delta_v + \mathbf{v}_m^T[k]\mathbf{v}_m[k]))} \tag{18}$$

and the modified energy term $\Delta\hat{E}_m[k]$ can be formulated as:

$$\Delta\hat{E}_m[k+1] = \begin{cases} \Delta\hat{E}_m[k+1] + \frac{1}{T_I}\left(\Delta E_m[k] - P_{\mathbf{f}_s}(\mathbf{v}_m[k])^T \delta(\mathbf{v}_m[k])\Delta\hat{E}_m[k]\right) \\ \quad \text{if } \Delta E_m[k] > P_{\mathbf{f}_s}(\mathbf{v}_m[k])^T \delta(\mathbf{v}_m[k])\Delta\hat{E}_m[k], \\ \quad \Delta E_m[k+1], \text{otherwise.} \end{cases} \tag{19}$$

4 Experimental Results

4.1 Description of the Experiments

To analyze the performances of the proposed bilateral control approach, two sets of experimental measurements were performed.

In the first experiment (*E1*) the teleoperation system was implemented using the facilities of two robotic laboratories situated in different European countries: Antal Bejczy Center for Intelligent Robotics, Óbuda University, Budapest, Hungary (slave side) and Robotics and Control Laboratory, Sapientia Hungarian University of Transylvania, Tirgu Mures, Romania (master side). The physical distance between these laboratories is about 500 km.

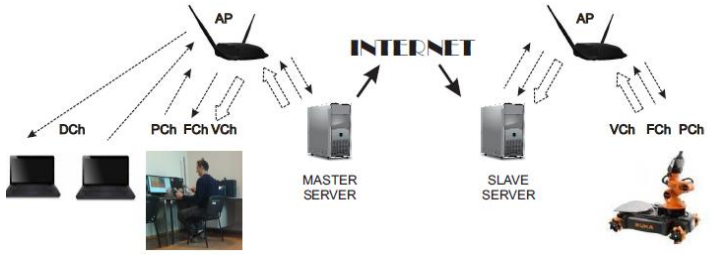


Figure 3
Experimental setup for measurements

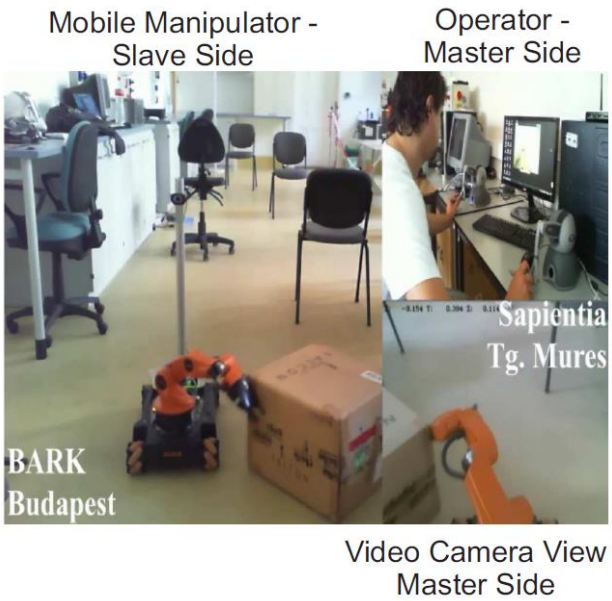


Figure 4
Teleoperation experiment

During the second experiment (*E2*) the master side was implemented at the Centre for Intelligent Machines, McGill University, Montreal, Quebec, Canada and the slave side was put in operation at the Robotics and Control Laboratory, Sapientia Hungarian University of Transylvania, Tirgu Mures, Romania. The physical distance between the master and the slave side during this transatlantic teleoperation experiment was about 7500 km.

In both experiments, a KUKA youBot mobile manipulator (KUKA Roboter GmbH, Germany) served as the slave robot. It has an omnidirectional mobile platform equipped with four Mecanum wheels; hence the longitudinal, lateral and angular velocities can separately be prescribed for the platform. A 5 DOF serial robot arm is placed on the manipulator and a gripper serves as the end effector for the slave robot. The control equipment of the arm provides position, velocity and force information about each joint separately. The robot used for teleoperation was also equipped with a USB video camera having 640X480 resolution; it renders video feedback about the motion of the slave robot.

During the experiment *E1* two PHANTOM Omni haptic devices (SensAble Technologies, Inc., USA) served at the master side to separately control the arm and the platform of the mobile manipulator. These haptic devices have 6 DOF and can provide force feedback on the first three joints. The first haptic device was applied to control the mobile platform. The first and third joints were used to prescribe the longitudinal and lateral velocity components, while the angular

velocity of the platform was set with the fifth joint. The platform was controlled by the human operator based on the video information provided by the camera mounted on the slave robot. The second haptic device was used to control the arm of the robot. The first 5 DOF of the second haptic device were used to control each joint of the arm. In the first three joints, the force feedback was implemented. Accordingly, the teleoperation system provides both haptic and visual information for the remote robot arm control. Using the buttons on the haptic devices the motions of the arm and the platform can separately be enabled and the gripper can be opened or closed by the operator.

In the experiment *E2* a Novint Falcon haptic device (Novint Technologies, Inc, USA) served to control both the arm and the platform of the mobile manipulator. The motion of the platform and the arm can be separately enabled with the buttons of the haptic device`s grip. For the distant control of the mobile platform the three DOFs of the haptic device were used to prescribe the longitudinal, traversal and angular velocities of the platform. In robot arm control mode, the same DOF's of the haptic device were applied to control the first three joints of the KUKA youBot's arm. During the bilateral teleoperation measurements, the force feedback was implemented for the first joint of the arm.

In Figure 3 the communication channels of the teleoperation system are presented. PCh (Position Channel) is used to transmit position and velocity information from the master side to the mobile slave. FCh (Force Channel) is applied to send the force data from the mobile slave to the master side. Through VCh (Video Channel) the video information is sent from the slave to the master. Other data channels can also be present in the communication medium applied to teleoperation (DCh – Disturbance Channel).

The software that was used for experimental measurements and implements the proposed bilateral teleoperation method is available on the Github.com web-based repository hosting service: https://github.com/TARC-Sapientia/TO.

A detailed description of the software can be found in [21].

4.2 Measurement Results - *E1*

The developed teleoperation system and bilateral control methods were tested through Internet-based teleoperation experiments. During the test scenario, the task of the human operator in Tirgu Mures was to navigate the KUKA youBot mobile manipulator in Budapest among obstacles to a target place. The mobile platform part was remotely controlled based on video information. During the motion, the mobile manipulator also had to push aside non-fixed obstacles by using its arm. These types of tasks are common in search and rescue applications. Fig. 4 presents the mobile manipulator, the operator and a frame of the video which was displayed during the teleoperation experiment.

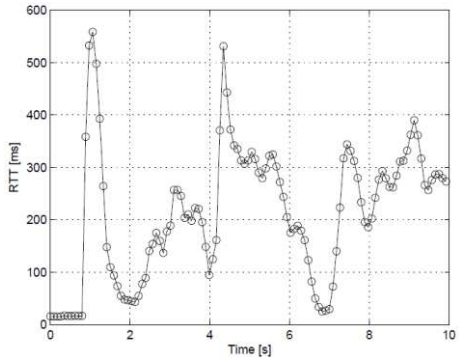


Figure 5
Typical Round Trip Time (RTT) evolution during teleoperation experiments

During the task execution, different bilateral teleoperation scenarios were tested. When the robot pushed the loaded box, the force exerted by the base joint of the arm was sent back to the human operator. The force felt by the human operator through the haptic device gave information to the operator about the magnitude of the friction between the box and the floor. For such cases, the reliable force reflection is important as through the haptic feedback the operator can decide whether the robot is capable of pushing aside the obstacle.

The Internet latency between Romania and Hungary is relatively small (approximately 30 ms). To test the teleoperation system in critical situations, during the experiments another UDP data flow in a bottleneck link of the master-slave communication channel was started (DCh). As bottleneck link the last hop wireless channel on the master side was considered, see Fig. 3. The communication congestion was achieved in this wireless network by activating an extra 5 MB/s UDP data flow. It manifested in increased delay and delay variation in the communication channel. A typical delay evolution during the teleoperation experiments is shown in Fig. 5. As the figure shows, the mean Round Trip Time values ($RTT = d_{ms} + d_{sm}$) were around 200 ms during the measurements, but delay peaks over 500 ms also appeared.

Both time domain passivity controllers presented in sections 2 and 3 were tested during the bilateral teleoperation experiment. In both cases a box, having the same weight, in the same location was pushed. In order to clear the box from its path the robot rotated its first joint approximately in the domain (-2 rad, 2 rad) and it was in direct contact with the box in the second part of the motion. In the case of the method presented in section 3 the following parameters were applied: δ_v = 0.01m/s and T_I = 10 s.

The experimental results in the bilateral control experiments are presented in the Figures 6 and 7. The measurements focused on the force feedback during the robot arm movement. Some similarities can be observed in both cases. First, the

bilateral teleoperation is stable, the difference between the master side output and slave side delayed input energy (E_m^{OUT}, E_s^{IN}) remains around zero; hence all three passivity controllers assure the stability of the teleoperation. On the other hand, a lag between the received reference position on the slave side (p_s) and the real joint position (p_r) can be observed. The lag is accentuated when there is contact between the robot arm and the environment.

In both cases the passivity controller activates when the operator starts retreating the robot arm. Fig. 6 shows the behavior of the original time domain passivity controller in the presence of delay variation. As this figure shows, the control signal has high-frequency components which temporarily degrade the quality of the force feedback. This can be avoided by applying the control law (17). The behavior of this control law is presented in the Fig. 7. As this figure shows, in both cases much better-quality force feedback was achieved.

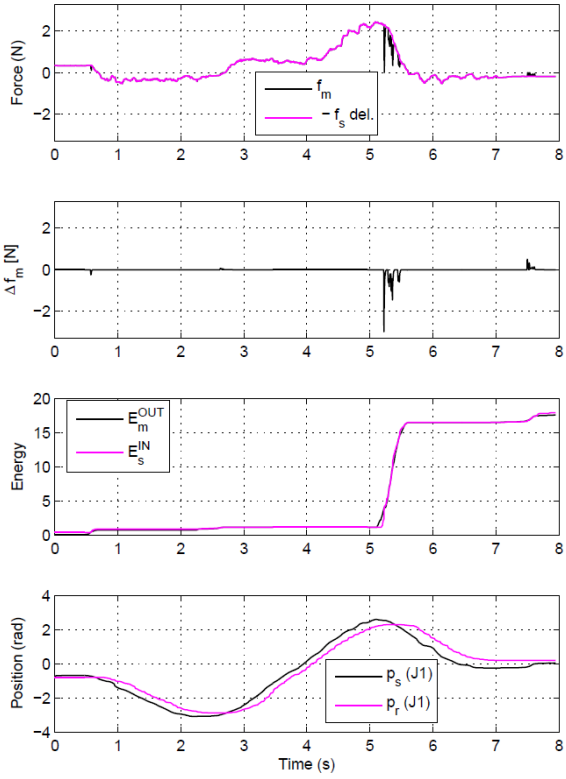


Figure 6

Bilateral control measurement (*E1*) with the original time domain passivity controller

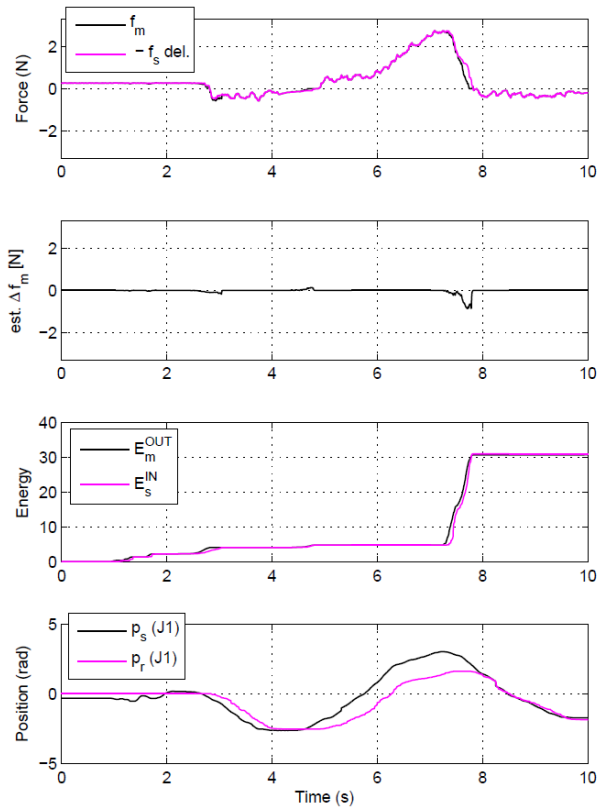


Figure 7
Bilateral control measurement (*E1*) with the proposed controller

4.2 Measurement Results - *E2*

A second set of transatlantic teleoperation measurements were also performed to show the repeatability of the proposed bilateral teleoperation scenario in other communication conditions.

A similar test scenario was elaborated as in the case of *E1*: on the master side (Canada) the human operator teleoperated the mobile manipulator, situated in Romania, among obstacles based on video and haptic information. Using its manipulator, the slave robot pushed aside non-fixed obstacles to advance toward its goal. The bilateral teleoperation was exploited during these events.

The Internet latency between Canada and Romania is approximately 135 ms. The higher delay and jitter values were achieved by starting an extra 5 MB/s UDP data flow in the wireless network that connected the mobile manipulator to the slave-side server. With this strategy, the average *RTT* was increased to 480 ms.

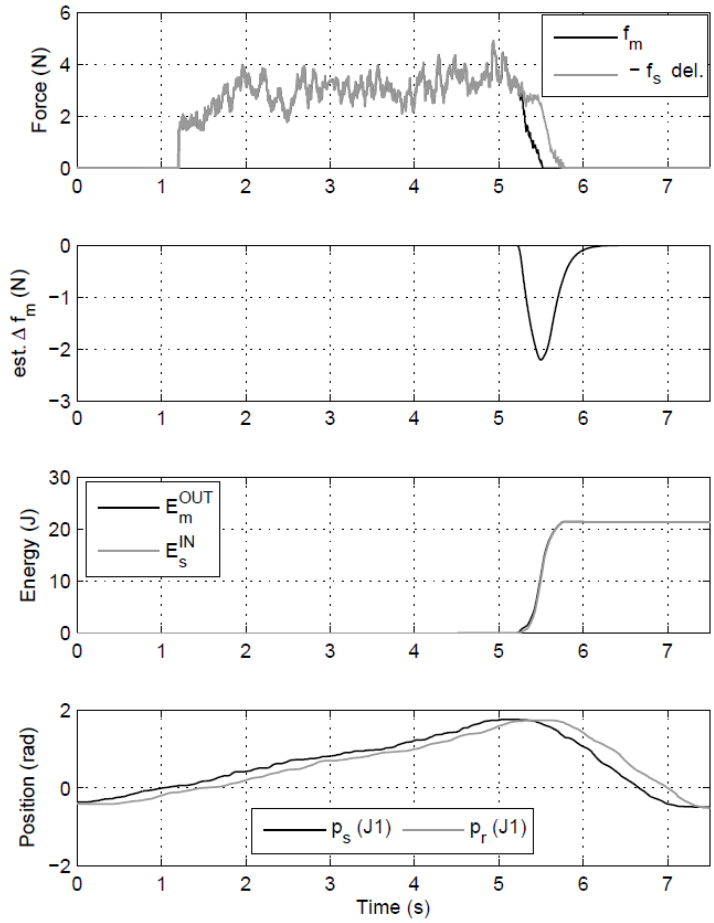


Figure 8
Bilateral control measurement (*E2*) with the proposed controller

During the teleoperation experiment the control method presented in Section 3 was applied with same implementation form and parametrization as in the case of *E1*. The experimental results are presented in Fig. 8. As it can be seen similar control performances are achieved as in the case of the experiment *E1*: with smooth control action, the stability of the teleoperation can be assured in the presence of large, time-varying communication delay. Similarly, as in the case of *E1*, p_s is the slave side position and p_r is the real joint position.

Conclusions

In this paper, a teleoperation system was presented for distant control of mobile manipulators. An enhanced passivity observer - passivity controller approach was introduced to assure stable and transparent bilateral teleoperation in the presence

of time-varying communication lag. The proposed algorithm is a modified time domain passivity controller that can ensure the stability of the teleoperation system with adequate force reflection, and with bounded control signal. The control assures smooth control actions even in the critical phases of the teleoperation, providing convenient operation conditions for the human operator.

Experimental measurements were performed using realistic teleoperation scenarios over mixed communication medium (Internet and wireless networks), including vision-based mobile platform navigation and bilateral control of the robot arm. The experimental measurements, performed over communication mediums that include both Internet and wireless network, show that the proposed bilateral control algorithms and software can assure both the stability and reliable force reflection in the bilateral teleoperation systems in different network communication conditions and with different commercial haptic devices.

Acknowledgement

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS UEFISCDI, project number PN-II-RU-TE-2011-3-0005. This work has been partially supported by ACMIT (Austrian Center for Medical Innovation and Technology), which is funded within the scope of the COMET (Competence Centers for Excellent Technologies) program of the Austrian Government. The research has been partially supported by the OTKA PD-114121 Hungarian Scientific Research Fund.

The research was partially supported by the Hungarian OTKA PD 116121 grant. This work has been supported by ACMIT (Austrian Center for Medical Innovation and Technology), which is funded within the scope of the COMET (Competence Centers for Excellent Technologies) program of the Austrian Government. T. Haidegger and P. Galambos are supported through the New National Excellence Program of the Ministry of Human Capacities. Partial support of this work comes from the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 project. This work was partially supported by the Hungarian EFOP-3.6.2-16-2017-00016 grant.

References

[1] Cen, Z., Mutka, M., Liu, Y., Goradia, A., Xi, N.: QoS management of supermedia enhanced teleoperation via overlay networks. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005, pp. 1630-1635

[2] Chawda, V., O'Malley, M.: Position synchronization in bilateral teleoperation under time varying communication delays. IEEE/ASME Transactions on Mechatronics 20(1), 2015, pp. 245-253

[3] Chopra, N., Berestesky, P., Spong, M.: Bilateral teleoperation over unreliable communication networks. IEEE Transactions on Control Systems Technology 16(2), 2008, pp. 304-313

[4] Ferre, M., Buss, M., Aracil, R., Melchiorri, C., Balaguer (Eds.), C.: Advances in Telerobotics, Springer, 2007

[5] Haller, P., Marton, L., Szanto, Z., Vajda, T.: Bilateral Teleoperation in the Presence of Jitter: Communication Performance Evaluation and Control, Springer, 2015, pp. 291-311

[6] Hannaford, B., Ryu, J. H.: Time-domain passivity control of haptic interfaces. IEEE Transactions on Robotics and Automation 18(1), 2002, pp. 1-10

[7] Hokajem, P. F., Spong, M. W.: Bilateral teleoperation: An historical survey. Automatica 42, 2006, pp. 2025-2057

[8] Huang, K., Lee, D.: Hybrid PD-based control framework for passive bilateral teleoperation over the Internet. In: Proc. of 18th IFAC World Congress, 2011, pp. 1064-1069

[9] Khan, Z. H., Genon-Catalot, D., Thiriet, J. M.: A co-design approach for bilateral teleoperation over hybrid network. In: Proc. of 18th Mediterranean Conference on Control & Automation, Marrakech, Morocco, 2010, pp. 286-291

[10] King, H., Hannaford, B., Kwok, K. W., Yang, G. Z., Griths, P., Okamura, A., Farkhatdinov, I., Ryu, J. H., Sankaranarayanan, G., Arikatla, V., Tadano, K., Kawashima, K., Peer, A., Schauss, T., Buss, M., Miller, L., Glozman, D., Rosen, J., Low, T.: Plugfest 2009: Global interoperability in telerobotics and telemedicine. In: IEEE International Conference on Robotics and Automation, 2010, pp. 1733-1738

[11] Lee, D., Huang, K.: Passive-set-position-modulation framework for interactive robotic systems. IEEE Transactions on Robotics 26(2), 2010, pp. 354-369

[12] Lee, D., Li, P. Y.: Passive bilateral control and tool dynamics rendering for nonlinear mechanical teleoperators. IEEE Transactions on Robotics 21(5), 2005, pp. 936-951

[13] Li, H., Kawashima, K.: Bilateral teleoperation with delayed force feedback using time domain passivity controller. Robotics and Computer-Integrated Manufacturing 37, 2016, pp. 188-196

[14] Marton, L., Szanto, Z., Vajda, T., Haller, P., Sandor, H., Szabo, T., Tamas, L.: Communication delay and jitter influence on bilateral teleoperation. In: Proc. of 22nd Mediterranean Conference on Control and Automation, Palermo, Italy, 2014, pp. 1171-1177

[15] Namerikawa, T.: Bilateral control with constant feedback gains for teleoperation with time varying delay. In: Proc. of 48th IEEE Conference on Decision and Control. Shanghai, China, 2009

[16] Nuno, E., Ortega, R., Barabanov, N., Basanez, L.: A globally stable PD controller for bilateral teleoperators. IEEE Transactions on Robotics 24(3), 2008, pp. 753-758

[17] Preusche, C., Hirzinger, G., Ryu, J. H., Hannaford, B.: Time domain passivity control for 6 degrees of freedom haptic displays. In: Proc. of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003, pp. 2944-2949

[18] Quang, H. V., Farkhatdinov, I., Ryu, J. H.: Passivity of delayed bilateral teleoperation of mobile robots with ambiguous causalities: Time domain passivity approach. In: Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 2635-2640

[19] Ryu, J. H., Artigas, J., Preusche, C.: A passive bilateral control scheme for a teleoperator with time-varying communication delay. Mechatronics 20, 2010, pp. 812-823

[20] Ryu, J. H., Kwon, D.S., Hannaford, B.: Stable teleoperation with time-domain passivity control. IEEE Transactions on Robotics and Automation 20(2), 2004, pp. 365-373

[21] Sandor, H., Szabo, T., Vajda, T., Haller, P., Szanto, Z., Marton, L.: Video supported bilateral teleoperation system: Design and implementation. In: IEEE 14th International Symposium on Computational Intelligence and Informatics, 2013, pp. 105-111

[24] Wirz, R., Marin, R., Claver, J.M., Ferre, M., Aracil, R., Fernandez, J.: End-to-end congestion control protocols for remote programming of robots, using heterogeneous networks: A comparative analysis. Robotics and Autonomous Systems 56(10), 2008, pp. 865-874

[25] U. Ahmad and Y. J. Pan, Switching time domain passivity control for multilateral teleoperation systems under time varying delays, IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, 2016, pp. 1429-1434

A Hardware Platform for Frequency Domain Spectroscopy and Frequency Response Analysis

Jan Tomlain, Ondrej Teren, Radek Sedlacek, Josef Vedral

Department of Measurement, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic
tomlajan@fel.cvut.cz, terenond@fel.cvut.cz, sedlacr@fel.cvut.cz, vedral@fel.cvut.cz

Abstract: This paper describes a development of a hardware platform suitable for the frequency domain spectroscopy (FDS) and frequency response analysis (FRA) applied for high-voltage machine diagnostics. The principle of the FDS and FRA methods widely used by maintenance technicians are briefly reviewed. Important technical requirements for the instruments utilizing these methods are stated. A design of a hardware platform with parameters fitting both of the method and engagement of the latest electronics components is discussed. In order for verification of the design platform, a basic implementation of a vector voltmeter is described. Long-term stability, phase shift sensitivity, and accuracy of the platform are evaluated.

Keywords: high-voltage machine diagnostics; dielectric spectroscopy; dielectric response analysis; frequency response analysis; non-invasive diagnostic; machine insulation; lock-in amplifier

1 Introduction

In the field of energetics, there are a number of high-voltage (HV) machines. As an example, power transformers and generators directly used in the production and the transmission of electricity, should be mentioned. The lifetime of the machines are expected to be decades. To achieve the expected lifetime it is necessary to plan the regular diagnostic examinations of these machines. Thus, the machine operator avoids an unexpected shutdown of the machine or the entire system [1]. In many cases, the power transformer can work in relatively good conditions even if they exceeded their operational lifetime [2]. Therefore, it is not economically effective to replace such machines solely because of exceeding their expected lifetime. This fact is another reason to develop a non-invasive assessment method of the machines.

HV machines windings and their insulation systems are constantly exposed to a combination of heat, electric and mechanical stresses [3], [4]. All of these factors have an influence on the state of the machines winding insulation. The state of winding insulation is one of the studied parameters that can detect an upcoming failure. There are few measurement methods used in winding insulation state assessment, such as: electric resistance measurement using DC current [5], partial discharge measurement [6], [7], dielectric response measurement [8], capacitance, and dissipation factor measurement measured at power grid frequencies, dielectric spectroscopy [9] and frequency response analysis (FRA) [10]. The stresses mentioned above and also aging and moisture generation induced by these stresses have an influence on the conduction and the polarization processes inside the insulation material.

This paper further deals with FDS and FRA diagnostic methods. Common hardware requirements for both methods are found, such as: (parameters of excitation signals of a device under test and sampling frequencies of the measurement of device response). Baseed on these requirements, a design of a hardware platform intended for the mentioned methods is described. The primary goal of this work is development of an instrument fully suitable for practical measurements utilizing the FDS and FRA methods. Another objective of this work, is also an effort to increase the accuracy of the measurement and improvement of parameters (frequency range, excitation voltage level, resolution, linearity, etc.) in comparison with other instruments available in the market.

2 Frequency Domain Spectroscopy

Frequency domain spectroscopy (FDS) is one of the methods suitable for studying polarization phenomena. Every kind of insulation material consists, at an atomic level, of negative and positive charges. The charges are balancing each other, and on the macroscopic level, they produce a neutral charge. When applying an external electrical field E the charges become oriented thus, formate dipoles. The positive charges are attracted to the negative pole and vice versa. This phenomenon is called the polarization.

FDS is a method that utilized an AC current excitation of studied material (transformer insulation). The capacitance and the dissipation factor are measured as a function of frequency of the excitation signal [11], that can be expressed using an equation

$$\mathbf{I}(\omega) = j\omega C(\omega)\mathbf{U}(\omega) \tag{1}$$

where $\mathbf{I}(\omega)$ is the measuring current, $C(\omega)$ is the capacitance of the measured material sample $\mathbf{U}(\omega)$ is the excitation voltage. The frequency range is usually between 1 mHz and 1 kHz, and a voltage level of hundreds of volts is needed [12].

A base block diagram of an FDS analyzer with a typical connection to a device under test (DUT) is shown in Fig. 1.

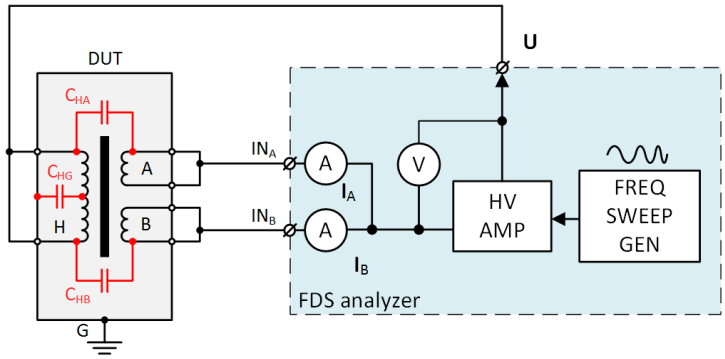


Figure 1
The block diagram of the dual channel FDS analyzer

The DUT is represented by a transformer with a primary winding H and two secondary windings A and B. There are also shown parasitic capacitances C_{HA}, C_{HB} and C_{HG} that can be measured using the FDS analyzer. The FDS analyzer consists of a frequency sweep generator (FSG), a high-voltage amplifier and two ammeters. The FSG generates a sinus waveform with an adjustable frequency. This signal is amplified using the high-voltage amplifier to an amplitude of hundreds of volts. A studied parasitic capacitance of the DUT is excited by the amplified signal. A current flowing through the studied capacitance is measured by the ammeters. In order to perform the computation of capacitance and dissipation factor, it is necessary to measure both the current amplitude and the current phase with respect to the generating signal. The measured parameters are calculated using the equation (1). A typical dissipation factor course measured using the FDS method is shown in Fig. 2.

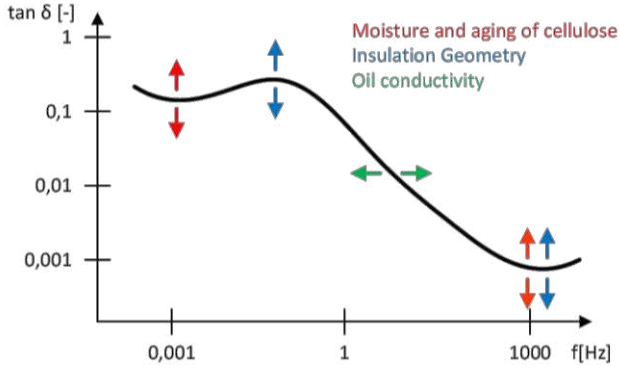


Figure 2
The typical FDS measurement result and factors affecting the dissipation factor, similar picture in [12]

3 Frequency Response Analysis

The motivation for the FRA measurements is very similar to the methods mentioned above. Many dielectric and mechanical failures in power devices like transformers are caused by mechanical changes in the winding structure. These defects, or displacements in the winding structure, may be the result of transportation damage occurring between the manufacturer and the installation location. Moreover, also powered devices by this particular transformer could cause defects. For instance, low impedance fault on these devices could be harmful [10]. Common problems are also natural effects of the insulation aging. Early detection of these displacements of a dielectric failure can reduce unplanned maintenance costs and provide the possibility of improving system reliability by preventing these failures. In addition, when the damage is discovered, repairs may be targeted to a specific place.

3.1 Theory

The main idea behind the FRA process is to measure the transfer function of the inspected subsystem in some frequency span. Inputs and outputs of the particular transfer function can be freely chosen to obtain various data.

Diagnostic result for the high voltage machinery FRA, is typically a relation between the impedance of the DUT as a function of the excitation frequency. The typical frequency range according to [10] is from 10 Hz up to 5 MHz. It has been verified that even minor displacements in the geometric structure of the large power transformer windings or changes in the dielectric parameters of the insulation system will lead to the relevant changes in the FRA fingerprint. This fact supports the results from simulations in [10] and [13]. To understand even better the relationship between the transfer function estimation and the transformer physical changes [13] provides a simplified equivalent schematic of the transformer with explanation of what value is changed, by what defect. A useful overview of these changes is documented in [14].

3.2 Swept Frequency Method

The first method, known as the swept frequency response analysis (SFRA) method, uses the fact that the impedance (or admittance) of a transformer winding varies with the frequency. A basic connection is shown in Fig. 3. The FRA analyzer could be substituted with a network analyzer for simplicity. DUT is excited with AC voltage with amplitude in a range of tens of volts, acc. to [10]. There is another voltage probe to measure actual voltage present on the DUT terminal. This connection eliminates any mistakes caused by testing voltage cable wiring. Current flown from neutral terminal of the DUT is the output measured

value. The input impedance of the network analyzer could be used to convert resulting current from the neutral terminal to the voltage as shown in Fig. 3. Despite the simple data analysis, there are several disadvantages of this method. At very low frequencies (in the range of some hundreds of Hz up to low kHz), these instruments tend to lack sufficient power to appreciably excite a large power transformer, due to the heavy inductive load presented by the steel core. This problem is described more deeply in [10]. The fixed sensing impedance of 50 Ω could be insufficient to convert the current signal in the whole frequency range properly. However, main advantages are simple hardware connection, easy data analysis processes and possible usage of instruments like network analyzers. On the other hand, disadvantages are in sensing impedance selection, cable length and connections, together with repeatability issue. A significant performance obstacle is an insufficient power to gain proper excitation of HV machines in the whole frequency range. SFRA method testing together with results is described in [15] and [16].

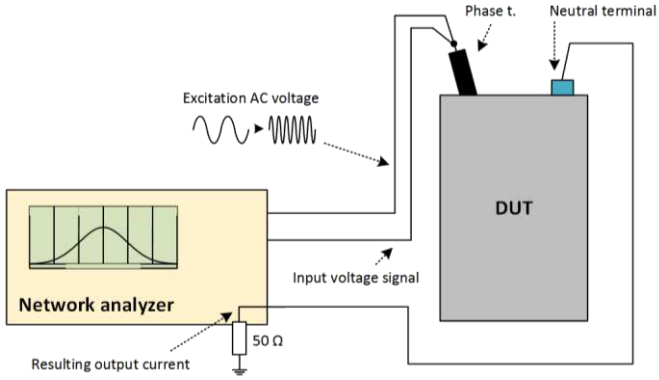


Figure 3
Typical test connection for a low voltage SFRA

3.2.1 Measuring Principle

The FRA instrument that is using the classical SFRA must measure two voltage signals, in which one corresponds to the voltage on the input phase terminal and the second one on the sensing impedance, which corresponds to the current-flown through DUT. The simplified electrical schematic is shown in Fig. 4.

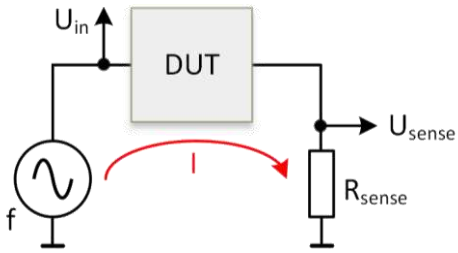


Figure 4
Measurement principle of DUT impedance

Impedance is then given by:

$$Z = \frac{U_{DUT}}{I} = R_{sense} \frac{U_{in} - U_{sense}}{U_{sense}} \tag{2}$$

3.3 Traditional Impulse Method

In contrast with, the previous method, the impulse method tries to compensate several imperfections of the swept mode. The basic principle is described in [10] and especially in [14]. Measurement is done in a frequency domain with single impulse transition response analysis. A test connection for the impulse measurement is shown in Fig. 5. A network analyzer is altered by the impulse generator with a high-speed digitalization unit. The main idea behind this measurement is to produce a single low voltage impulse to one terminal of the DUT and measure particular response to this impulse on the neutral terminal. Similarly, the current on the neutral terminal should be properly converted to the voltage. There is a possibility to use another value of the sensing impedance in contrast with internal one in the network analyzer. Once the time domain record of the voltage (input) and current (output) are recorded, the transfer function is calculated as the Fast Fourier Transform (FFT) of the output divided by the FFT of the input. Unfortunately, this method has also its limitation. There are several practical problems which cause difficulties with the repeatable results together with the accuracy aspect. Present noise in the signal has a negative effect on the final FFT result. Easiest elimination is to apply averaging of more than one result. This is one of the recommendations in [10]. A good level of repeatability is limited by another two factors. Test impulse should be the same for each single measurement of the averaged set. The second, is the quality and performance of the digitalization unit. The Impulse FRA method testing together with results is described in [17] and [18].

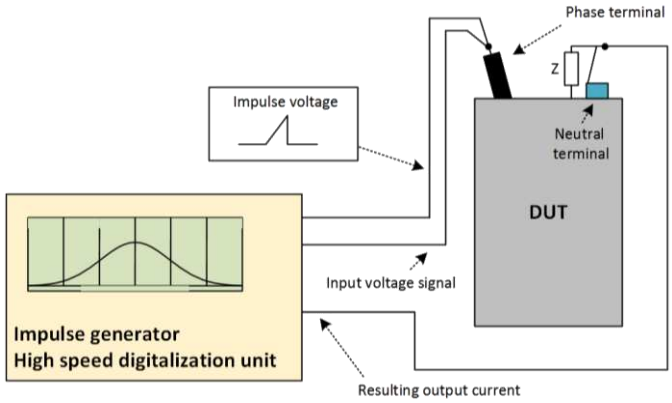


Figure 5
Typical test connection for a low voltage SFRA

4 Designed HW Platform

The FDS and FRA methods described in Chapters 2 and 3 do have some common attributes. Both of the methods require a generation of an excitation signal and a measurement of the DUT response. The measurement is based on sampling the response signal, but the signal processing itself is different for both methods. However, there is a possibility to integrate both methods and implements of them into one platform.

Hardware for the whole diagnostic process consists of three main parts, see Fig. 6. A core part of the digital signal analysis is an FPGA board, a data conversion and acquisition is done on an analog daughterboard. A coupling front-end board provides a power front-end to amplify signals generated by the daughter board. The coupling front-end is designed to operate with excitation signal with a voltage of 30 Vpp and fulfills requirements of the excitation signal in case of FRA method. Thus, then a chain of FPGA board, daughter board and coupling board forms whole FRA instrument. In the case of the FDS method, an external HV amplifier is needed that is not discussed in this paper.

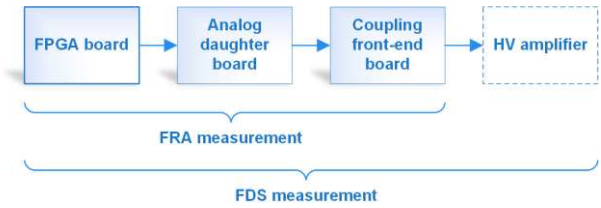


Figure 6
The block diagram of the platform

The designed system will use frequency swept method, which is described in 3.2. According to this method, designed HW should fulfill the possibility to work in frequency spectrum from tens of Hz up to tens of MHz. In addition, communication and interfacing the instrument must be robust with a usage of modern communication interfaces like Ethernet.

4.1 FPGA Board

Core parts of the digital subsystems are placed in the field-programmable gate array (FPGA) board. To obtain an easy implementation of real-time parallel tasks and large scalability the FPGA architecture was chosen. Altera Arria V GX device in FBGA672 case provides a reasonable ratio between requested functionality and price. Two memory interface for DDR3, a large amount of internal RAM and high-speed GPIO interfaces are the essential parts. In addition, the user is able to implement a high-speed System-on-the-chip (SoC) systems using Nios 2 Gen2 soft processor core, in this device family. Control between the FPGA board and host/slaves systems is ensured via several types of communication standards. Ethernet 10/100 Mbit/s interface is provided by an external TCP hard-wired stack manufactured by Wiznet company. The W5500 chip provides up to 8 parallel full-featured TCP/UDP stacks. The internal MAC possibility of the Altera device is not used because of licensing policies. This solution provides comparable data throughput and much fewer computation requirements to the host SoC system inside FPGA. The full-duplex RS485 interface is implemented for an industrial standard communication via twisted-pair buses. This interface is also isolated. For convenient PC to HW communication the FPGA board is equipped with USB interface. Internal RAM of 8 Mbits could be expanded with two DDR3 chips each 64 MB. High Speed Mezzanine Card (HSMC) 160 pin connector is designed on the edge of the board for connection with the daughterboard. This interface consists of 19 differential gigabit links, 32 LVDS differential connections, 20 LVTTL single GPIO and the power supply network. Character LCD display and JTAG connector could be used for debugging or simple results visualization. Power circuits consist of complex power supply solution for whole FPGA systems with the proper voltage ramping and powering sequence. The simplified block diagram is in Fig. 7.

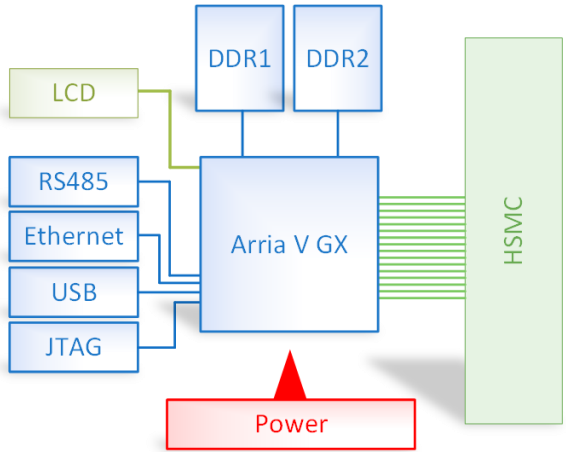


Figure 7
The block diagram of the motherboard

4.2 Daughter Board

A board equipped with an analog to digital (ADC) and a digital to analog (DAC) converters was designed. The board is mainly intended for the measurement method described in Chapter 2. A block diagram of the board is shown in Fig. 8.

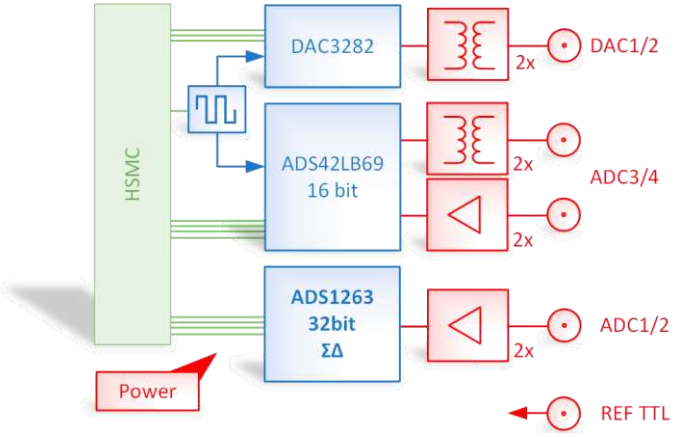


Figure 8
The block diagram of the designed analog front-end board

The board consists of one dual channel DAC type DAC3282 (16-bit, 625 Msps). Its output is led to an SMA connector. The board is also equipped with two dual channel ADC. The first ADC is type ADS42LB69 with 16-bit resolution and 250 Msps sampling rate. An input signal led from SMA connector is coupled both AC

(transformers) and DC (operational amplifier) ways. This topology allows using the ADC in a wide frequency range of the input signals. The second ADC used was, type dual channel ADS1263 with a 32-bit resolution, maximum sampling rate 38 ksps and a sigma-delta architecture. In comparison with the first ADC, it provides better resolution and accuracy in the range of low frequencies. Clock impulses for DAC and ADC are generated using a circuit Si5338. The Si5338 is a programmable any-output any-frequency quad clock generator. The board also contains all necessary power supply circuits. The power supply circuits provide all voltage levels for: ADC, DAC and signal preconditioning for analog circuits. An REFTTL signal led into a high-speed mezzanine connector HSMC is proposed to a sensing of a square wave reference signal. This can be used for internal phase-locked loop synchronization.

4.3 Coupling Front-end Board

One of the important instrument parts is an analog interface board between the analog daughterboard described in Chapter 4.2 and the particular DUT. The proposed structure of this board is shown in Fig. 9.

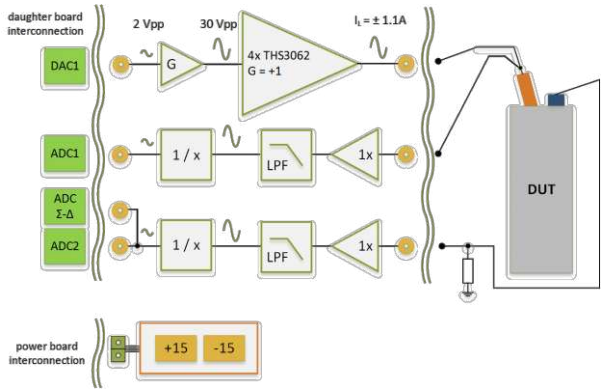


Figure 9
Proposed design of excitation FRA board

As the output buffer, the THS3062 from Texas Instruments will be used. Its frequency range in the gain range of about 1 is up to 300 MHz, which is fully capable of the proposed design. It can supply up to ± 145 mA per channel. Interconnection of eight parallel channels will produce more than 1.1 A to the load in the form of the DUT. The power output capability is not used continuously because of the relationship between the DUT impedance and the particular frequency.

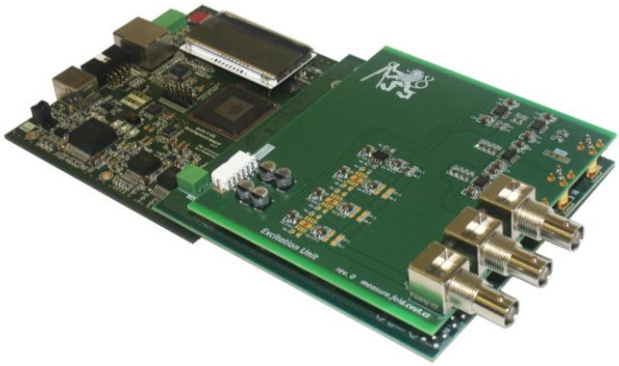


Figure 10
Photograph of the designed platform consisted of all three designed board

Input signals should be low-pass filtered to eliminate high-frequency noise and amplitude of the signal will be adapted to the low-voltage signal for ADC inputs on the daughterboard. A photograph of the whole assembled platform is shown in Fig. 10.

5 Design of FPGA Software Core

Both of the methods described in Chapters 2 and 3 require measurement of amplitude and phase of harmonic signals. Thus, a basic vector voltmeter implementation was designed to prove a suitability of the designed platform for the method described in Chapter 4. The implementation is running on the FPGA board described in Chapter 4.1. A principle of used algorithm is shown in Fig. 11 and it is also well-known as a lock-in amplifier (LIA). The LIA is one the commonly used techniques used to evaluate amplitude and phase of the harmonic signals. The algorithm consists of two multipliers, two low-pass filters, and a phase shifter. U_S is an input sinus wave signal and U_R is a reference sinus wave signal.

Then the output signals U_X and U_Y can be expressed as

$$U_x = \frac{1}{2}U_S U_R \cos[(\omega_s - \omega_r)t + \theta_s - \theta_R] - \frac{1}{2}U_S U_R \cos[(\omega_s - \omega_R)t + \theta_s + \theta_R] \quad (3)$$

$$U_x = \frac{1}{2}U_S U_R \cos[(\omega_s - \omega_r)t + \theta_s - \theta_R] + \frac{1}{2}U_S U_R \cos[(\omega_S - \omega_R)t + \theta_S + \theta_R] \quad (4)$$

that indicates that the output signal U_X is proportional to a real part of the input signal and the output signal U_Y is proportional to an imaginary part of the input signal.

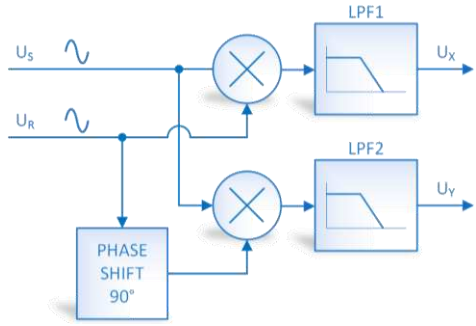


Figure 11
The principle of the lock-in amplifier

A complete overview of the designed FPGA core is in Fig. 12. The parallel tasks connected with ADC readout together with mathematical computations are implemented in the VHDL. While the data representation and communication interface is implemented inside NIOS II processor.

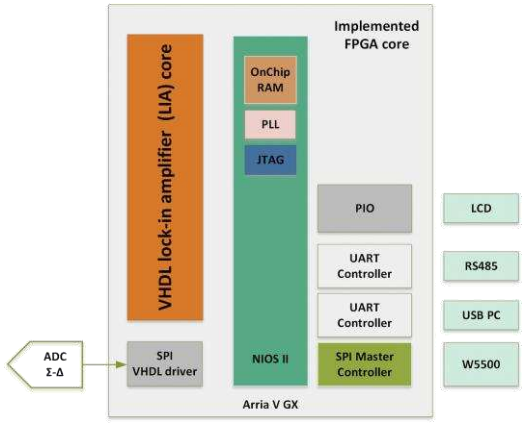


Figure 12
Block diagram of FPGA core

A block diagram of a LIA VHDL implementation suitable for the FPGA board is shown in Fig. 13. An ADS1263 VHDL driver serves an interface between ADC and FPGA circuit. The driver provides continuous a reading of the samples, and it also determines the sampling frequency. Sampled data are continuously provided in the bus data. The sampled data are multiplied in multiplier MUL32. The second input of the multiplier is fed by an all digital phase-locked loop PLL. The PLL generates a sine wave and a cosine wave digital waveform with a phase shift respected to the REF signal. The filtration of the frequency $\omega_S + \omega_R$ is provided by a PERAVG and an MOVAVG filter. The PERAVG is designed as an average of one period of the input signal, and MOVAVG is designed as a moving average

filter. The filtered signal is led into a NIOS processor block that provides a necessary computation logic.

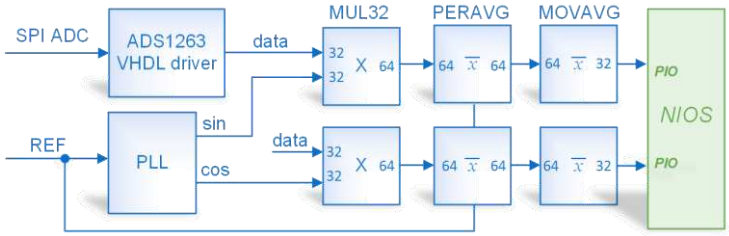


Figure 13
The VHDL implementation of the LIA

6 HW Platform Verification

A few experiments were made to verify the designed platform. For both methods (FDS and FRA) the key result depends on the precise measurement of amplitude and phase of two input signals with reference to excitation signal. Therefore, the most important part of the whole FDS and FRA analyser consists in quality of LIA implementation. The first experiment, was focused on the long-term stability of used ADC, the second experiment was focused on the sensitivity of a phase shift measurement, and the third experiment was focused on the accuracy of the phase shift measurement. The block diagram of the first and the second experiment is shown in Fig. 14.

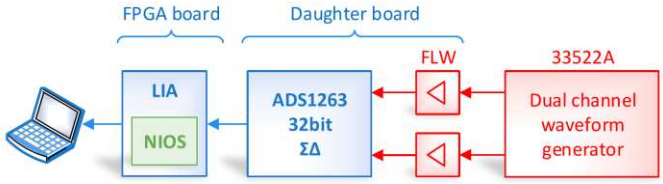


Figure 14
The block diagram of the stability and phase sensitivity experiment

In Fig. 15 is plotted data measured by the LIA algorithm (described in Chapter 5) with a stable sinus waveform, with a frequency of 50 Hz led into the input of the front-end board. During the two hours test, no drift was observed. The root mean square error was evaluated as 1.5 μV_{RMS}.

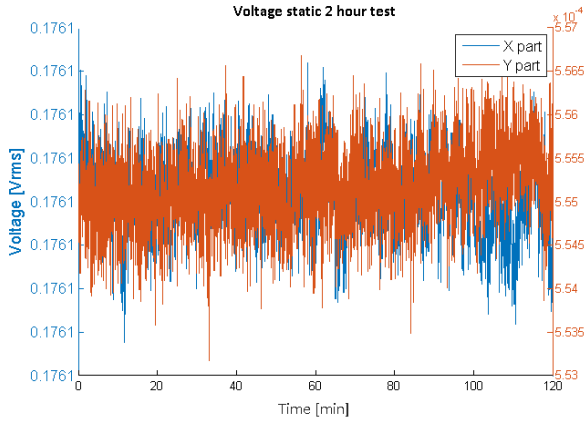


Figure 15
The results of the phase sensitivity experiment

The second experiment was designed to prove the phase shift sensitivity close to 0.001˚ that leads to the tangent delta of $1.75 \cdot 10^{-5}$. The tested signal was generated using an arbitrary waveform generator Agilent 33522A. A minimal phase shift that can be generated using this generator in the arbitrary mode at the frequency of 50 Hz is of 0.00108˚. The sensitivity of the phase shift change of 0.00108˚ is visible from the measured data shown in Fig. 16. It is evident that the designed measuring system is able to detect at least the phase shift changes of 0.00108˚.

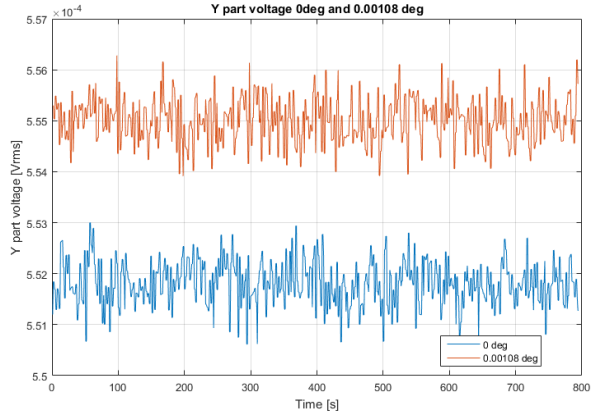


Figure 16
The results of the phase sensitivity experiment

The third experiment's goal to prove the accuracy of the measurement of the phase shift. A block diagram of the experiment is shown in Fig. 17. The front-end was excited by sin waveform with the frequency of 50 Hz, and internal phase

shifter was set at different angles. Measured results are stated in Table 1, where U_{IN} is an RMS value of input voltage, φ is a phase shift set into the internal PLL phase shifter, θ is phase computed from output values of the implemented LIA algorithm and ε is an absolute error with respect to the set internal phase shift.

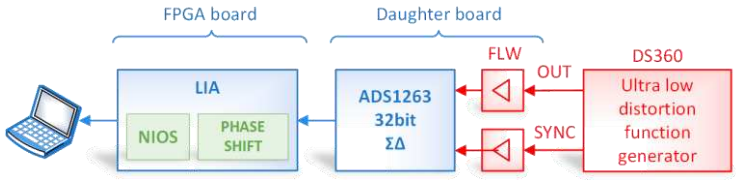


Figure 17
The measurement of phase shift accuracy

Table 1
The results of the angle measurement accuracy

U_{IN}[V]	φ [rad]	θ [rad]	ε [rad]
0.17966	0 Π	$-2.55 \cdot 10^{-5}$	$2.55 \cdot 10^{-5}$
0.17966	1/8 Π	0.39267	$2.91 \cdot 10^{-5}$
0.17966	1/4 Π	0.78537	$2.82 \cdot 10^{-5}$
0.17966	3/8 Π	1.17807	$2.72 \cdot 10^{-5}$
0.17966	1/2 Π	1.57077	$2.63 \cdot 10^{-5}$
0.35934	0 Π	$1.6 \cdot 10^{-5}$	$1.6 \cdot 10^{-5}$
0.35934	1/8 Π	0.39272	$2.09 \cdot 10^{-5}$
0.35934	1/4 Π	0.78540	$1.84 \cdot 10^{-5}$
0.35934	3/8 Π	1.17811	$1.28 \cdot 10^{-5}$
0.35934	1/2 Π	1.57078	$1.63 \cdot 10^{-5}$

Conclusions

This paper describes a design of an HW platform intended for the frequency response analysis and the frequency domain spectroscopy applied for HV machines diagnostics. Briefly mentioned is the principle of the used diagnostic methods and the requirements for an instrument implementing both methods are pointed out. The motherboard provides necessary computation to implement an algorithm of the FDS and FRA methods. To verify a functionality of the platform a basic lock-in amplifier implementation was designed. It was observed that the daughter board provides time stable results and the instrument is able to use the measurement of the dissipation factor with a resolution of $1.75 \cdot 10^{-5}$ at the frequency of 50 Hz necessary for the FDS method. It was also proved an accuracy of $3 \cdot 10^{-5}$ rad in measured angle, although without any sophisticated calibration.

To complete the FDS analyzer it is necessary to solve the output high-voltage amplifier (see Chapter 2). The next work will be oriented toward the testing of a commercially available amplifier. After verification of the amplifier in a practical

FDS measurement, the integration of the amplifier into FDS analyzer will be considered. Also, a calibration method of the analyzer is supposed to be developed.

Acknowledgement

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS15/213/OHK3/3T/13.

References

[1] J. S. Simons, "Diagnostic testing of high-voltage machine insulation. a review of ten years' experience in the field," IEE Proceedings B - Electric Power Applications, Vol. 127, No. 3, pp. 139-154, May 1980

[2] G. K. Supramaniam, Z. F. Hussien, and M. A. T, "Application of frequency domain spectroscopy (fds) in assessing dryness and ageing state of transformer insulation systems," in Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International, Dec 2008, pp. 55-61

[3] J. P. van Bolhuis, E. Gulski, and J. J. Smit, "Monitoring and diagnostic of transformer solid insulation," IEEE Transactions on Power Delivery, Vol. 17, No. 2, pp. 528-536, Apr 2002

[4] T. K. Saha, M. Darveniza, D. J. T. Hill, and T. T. Le, "Electrical and chemical diagnostics of transformers insulation. b. accelerated aged insulation samples," IEEE Transactions on Power Delivery, Vol. 12, No. 4, pp. 1555-1561, Oct 1997

[5] M. S. Naidu, V. Kamaraju, and M. S. Naidu, High Voltage Engineering. McGraw-Hill Professional Publishing, 1995

[6] M. Farahani, E. Gockenbach, H. Borsi, and M. Kaufhold, "A method for the evaluation of insulation systems for high voltage rotating machines," in Properties and Applications of Dielectric Materials, 2003. Proceedings of the 7th International Conference on, Vol. 3, June 2003, pp. 1108-1111

[7] R. Schwarz and M. Muhr, "Diagnostic methods for transformers," in Condition Monitoring and Diagnosis, 2008. CMD 2008. International Conference on, April 2008, pp. 974-977

[8] B. H. Ward, "A survey of new techniques in insulation monitoring of power transformers," IEEE Electrical Insulation Magazine, Vol. 17, No. 3, pp. 16-23, May 2001

[9] M. Farahani, H. Borsi, and E. Gockenbach, "Dielectric spectroscopy in time and frequency domain on insulation system of high voltage rotating machines," in Solid Dielectrics, 2004. ICSD 2004. Proceedings of the 2004 IEEE International Conference on, Vol. 1, July 2004, pp. 60-63

[10] J. Britton. (2004) Transformer maintenance and diagnostics using frequency response analysis. [Online]. Available: http://www.electricenergyonline.com

[11] J. Abhishek and P. Aaradhi, "Dielectric diagnosis of ehv current transformer using frequency domain spectroscopy (fds) & polarization and depolarization current (pdc)," International Journal of Scientific & Engineering Research, Nov 2012

[12] OMICRON. (2016) Dielectric response analysis and moisture in oil-paper dielectric. [Online]. Available: https://www.omicronenergy.com/fileadmin/user upload/pdf/literature/DIRANA-Brochure-ENU.pdf

[13] A. Abu-siada, N. Hashemnia, S. Islam, and M. S. A. Masoum, "Impact of transformer model parameters variation on fra signature," in Universities Power Engineering Conference (AUPEC), 2012 22nd Australasian, Sept 2012, pp. 1-5

[14] Z. Wang, J. Li, and D. M. Sofian, "Interpretation of transformer FRA responses 2014; part i: Influence of winding structure," IEEE Transactions on Power Delivery, Vol. 24, No. 2, pp. 703-710, April 2009

[15] U. Sharma, S. Chatterjee, and K. Bhuyan, "Development of reference SFRA plot of transformer at design stage using high frequency modelling," in Power and Energy in NERIST (ICPEN), 2012 1st International Conference on, Dec 2012, pp. 1-4

[16] S. S. Patil and S. E. Chaudhari, "An attempt to investigate the transformer failure by using DGA and SFRA analysis," in Properties and Applications of Dielectric Materials (ICPADM), 2012 IEEE 10th International Conference on the, July 2012, pp. 1-4

[17] P. T. M. Vaessen and E. Hanique, "A new frequency response analysis method for power transformers," IEEE Transactions on Power Delivery, Vol. 7, No. 1, pp. 384-391, Jan 1992

[18] L. Coffeen, J. Britton, and J. Rickmann, "A new technique to detect winding displacements in power transformers using frequency response analysis," in Power Tech Conference Proceedings, 2003 IEEE Bologna, Vol. 2, June 2003, p. 7

Factors Determining the Development of Business Relationships in the Advertising Market

Erika Seres-Huszárik, László Józsa, Zsuzsanna Tóth

J. Selye University, Bratislavská cesta 3322, 945 01 Komárno, Slovakia
huszarike@ujs.sk, jozsal@ujs.sk, tothz@ujs.sk

Abstract: Factors affecting the success of business relationships for a long while have been in the focus of those specialists and professionals who are interested in the topic over the last two decades have caused significant changes in everyday life of participants in the economy. Over the past decades the concept of relationship marketing has become well-known and widely spread in business life. It is a well-known fact that the cost of acquiring a new customer is much higher than the cost of keeping an existing one. Participants of the economy define relationship marketing as a strategy to achieve customer loyalty. During the initial phase of the relationship, the company aims to establish a strong relationship with the customer, to win indifferent customers and make them loyal. Based on the above-mentioned factors, we found it important to asses the factors affecting the agency-client relationship. Based on several empirical studies, in our research we tried to answer the question whether it can be stated on the basis of the unanimous opinion of the participants of the Slovak advertising market that during the development of business relationships interpersonal relationships play the most important role ((hypothesis testing). Results of our research will help us understand the development of the agency-client relationships, and provide practical results for the advertising industry.

Keywords: advertising agency; client; relationship

1 Introduction

Factors affecting the success of business relationships, for a long while have been in the focus of those specialists and professionals who are interested in the topic, since the last two decades have caused significant changes in everyday life of participants in the economy [1]. Adaptation of the traditional and well-known McCarthy´s [2] marketing mix perfectly fitted into the influence of North America's mass production in the 50s and 60s, just as it fits into the well-organized sale system and into the mass media affected by the dominant

participants in the market. Grönroos [3] and Grayon-Ambler [4] also criticized the applicability of the 4P model. Both authors agree that marketing mix model is only a list of components by which processes cannot be traced. According to Gummesson [5], the 4Ps of marketing define the components of marketing from the aspect of manufacturers, and it deals less with the demand-side issues such as handling complaints and billing. According to Baker et al. [6], while applying the 4Ps of marketing a company makes decisions with the customers rather than for the customers. Based on the facts outlined above, it is essential that today's managers really focused on building good business relationships with customers. Termination of a relationship means significant time, money and energy costs [7].

1.1 Importance of Relationship Marketing

Nowadays, relationship marketing (RM) is a very common area of research in the field of marketing, as it provides competition tools in many industries by which we can acquire customer loyalty. The most important goal of RM is to build a strong relationship with customers just as to gain sympathy and later commitment of those customers who dislike our business [8]. In business life, RM raises stability, reduces uncertainty and blocks competitors, as a stable consumer base is formed by its customers [9].

The concept was first used by Berry in 1983 at a conference organized by the American Marketing Association [10]. According to Berry, the key factor that influences customer acquisition and retention is the high quality of services offered by the company.

In his research he also pointed out the fact that it is much more expensive to gain a new customer than to retain an existing client base, so companies in order to avoid possible disappointments need to pay much more attention to meet the needs of their existing customers. For these reasons, a new paradigm was needed in marketing [11], [12], [13], but some of the researchers did not accept it, although according to Harker and Egan [14], it is beyond reason to deny that it is an existing phenomenon. In the course of the paradigm shift, the traditional marketing transaction model is replaced by the model of RM. Francová and Oreský [15] draw attention to the fact that the interpretation of RM becomes much more tangible when we compare the specialty with the transaction-oriented approach. The name already refers to the fact that the main point of the transaction-oriented marketing approach is the process of intra-company transactions. In other words, the transaction-oriented approach is based on the needs of the company, which is more important than satisfying customer needs. On the contrary, the essence of RM is to build an active, continuously developing, long-term relationship between the company and its customers.

1.2 Interpretation of the Dimensions of RM

In marketing literature, we find references to some dimensions of RM such as commitment, trust, cooperation, etc., which are essential in the development of marketing strategies. In corporate practice, the application of RM will provide a competitive advantage [16]. It is a tool for retaining customers in the relationship, for helping survive the potential failure of services [17] and for creating a more inclusive customer base [18]. Nowadays, RM is undergoing a major transformation within the holistic marketing approach in order to help customer acquisition and retention by sharing and fulfilling promises [19]. In the work of researchers in the field of RM, 23 dimensions of RM have been identified. Sharma and Patterson [20] examined customers' views on alternatives. They focused their research on the process of selecting the available and advantageous business offers by clients. During the establishment of the business relationship, the bond between the parties is a very important [21] dimension which, as a result of a psychological process, means a strong, beneficial link for both parties. This is followed by commitment to the relationship. Gounaris [21] considers commitment as a desire for continuity, which means that business partners share resources. In order to develop commitment, it is necessary to have a smooth flow of information between the parties [22] in both formal and non-formal channels. In customer RM, competence of the staff [23], that is the ability to be able and willing to solve problems, conflict management [24], co-operative attitudes [25] as well as empathy, is necessary in problem areas.

From the management side, the application of RM requires co-ordination skills of managers, i.e. joint efforts, strength to be able to control and influence the behavior of the other party [26]. Regarding the subject of exchange, the preparation of personalized products/services [27] and their quality are important. If the customer positively evaluates the proportion of the achieved profit and investments, then the customer is satisfied [28] and then they depend on each other [29] and evaluate the benefits of the relationship [30]. In the assessment of the resources invested in the relationship [31], common goals [32], reciprocity [30], common values [22], as well as trust emerge [33].

1.3 Applying RM in Corporate Practice

RM should be part of corporate strategies since for a long-term customer retention great efforts are needed. Today's corporate practice is characterized by the fact that companies have recognized the importance of RM and are trying to introduce and apply it when dealing with their partners (trading partners, suppliers). Experience has shown that, thanks to RM programs, customer base grows, sales volume and thus profitability of the company increases. Customer loyalty programs are the main tools of RM efforts. They have a critical role in acquiring, retaining and developing relationships between the parties [34], [35]. The main

indicator of the application of the program is that the attention of the companies is aimed at the customer and not at the product. Creation of databases and call centers are the first steps when implementing the RM programs.

In the light of the above we can say that the practice of RM is a widely accepted practice in some industries due to the increase in market competition. Nowadays, most companies use RM tools to sell their products. They recognized that if they want to sell a certain product (e.g. financial products), it is essential to gain customer trust and to build a strong relationship with the customer. Marketing environment has undergone significant changes thus companies should aim to build trust with their customers, thereby protecting the relationship with their competitors' offers. So it is definitely necessary to create a marketing strategy, focusing on suppliers, partners and clients in order to maintain the business relationship. In their research, Liang et al. [36] found that in corporate development strategies, investing in customer relations can serve as a basis for creating customer value. Customer value means a sustainable competitive advantage resulting in solid financial performance.

In order to adapt RM to corporate practice, we need to automate business processes with different technology solutions and information databases. Of course, development of IT technology, development of the Internet and the widespread applicability of its possibilities were crucial. In RM, technology can support the collection, coordination and analysis of market and other data, which greatly contributes to the development of a good marketing strategy. The next part deals with the importance of using some tools resulting from the connection between RM and technology.

Thanks to e-CRM, companies can gain a lot of benefits as the application facilitates the flow of information within and outside the organization. According to Kotorov [37], e-CRM is an information and communication application that can increase the range of customer services. According to Kelley et al. [38], e-CRM includes marketing activities, tools and techniques (web site, e-mail, data acquisition, data mining, warehousing) that are implemented over the Internet with the aim of building long-term customer relationships. According to Da-wei [39], the ultimate goal of CRM is to build a satisfied and loyal customer base. He argues that all online businesses should use e-CRM as an integrated part of business processes in order to increase e-loyalty. Security issues arise, as partners share information and databases for more efficient business transactions, which may result in unauthorized access to data, resulting in loss of trust and it may mean the termination of the relationship [40]. Using technology, it has become possible for companies to approach their potential customers through online systems to provide information and feedback on their products. In current business practice, it is quite common that companies use online feedback or products/services evaluation by customers. To do this, a user-friendly system is necessary that will make it easy for the customer to contact the seller. The application of real-time CRM is an important technology tool in the field of RM,

with the aim of responding to the events when they actually happen. This is a practical application in the field of event marketing, automation of processes, just in time provisions and business intelligence. It is about quick information gathering, faster business decisions and up-to-date monitoring of changes in the market environment [41].

The adaptation of technology opportunities in the B2B sector was first analyzed by O'Callaghan et al. [42]. The internet and its technological advances have significantly increased marketing opportunities and transformed relationships between companies and their customers. Patterson et al. [43] identified seven factors that affect a company in B2B markets to apply technological opportunities: supply chain strategy integration, pressure of supply chain members, transaction environment, environmental uncertainties, organizational performance, organizational size and structure.

1.4 Impact of the Application of RM on Performance and Market Development

We can find a number of studies that examine the relationship between RM and company performance. Authors agree that RM's system of tools has an impact on corporate performance, as it helps increase trust and commitment, as well as to build a strong emotional bond between the parties. RM has a positive impact on business performance [44], as it provides the company with strong and stable customer relationships that provide proper word-of-mouth advertising, reduce employee training costs, fluctuations, and these result in higher profits.

There is also a positive relationship between the development of corporate brand and performance. Close relationship strategy implies that customers are much more active when establishing and maintaining a relationship, and are more satisfied and loyal. RM provides an adequate framework to understand how the company can achieve customer loyalty and build a corporate brand focusing on important aspects of the relationship. Regarding the strategic part, corporate branding has an important role in building relationships with shareholders, partners, customers and employees. Strategy, that is based on getting to know the partners, is a relatively new line of research in the field of corporate branding [45].

Based on the above, we can conclude that, as a result of business growth, the company's market share increases, customer loyalty, satisfaction and confidence increase, resulting in profitability and market development. Market development of the company is, therefore, influenced by a number of factors. First of all, it is worth mentioning the matter of loyalty. Customer loyalty cannot be built with promotions and marketing campaigns, but rather with the value that the product/service provides. Customer loyalty was defined by Oliver [46] as "a deep commitment to a product/service that persuades the customer to repurchase the same brand or brand group. Despite all the efforts of the company, behavior of the

customers may change and we may lose their loyalty." According to Palmatier et al. [47], customer loyalty determines company profitability and growth. In their later [48] studies, they found a positive relationship between the value gained by the customer and loyalty to the company. Homburg et al. [49] and Huang [50] also found a strong positive correlation between customer orientation and loyalty to the company. Customer satisfaction is a further factor, which is a response after an emotional purchase that is realized as a result of the comparison of expected and actual performance [51]. Satisfaction as a factor has a unique role in the service sector. Bennett and Rundle-Thiele [51] in their four studies on Australian banks point out that 23 to 32% of customers are dissatisfied with banking services. Nevertheless, the banks surveyed are listed on the TOP6 list of state-owned companies in Australia regarding the profit. This shows that some of the dissatisfied customers may remain loyal.

Business trust is when we are convinced of the trustworthiness and honesty of the other party [52]. Trust is a prerequisite for increased engagement. To define trust, we can find many definitions in the literature of RM. Nykanen, Jarvensivu and Moller [53] stated that confidence can clearly reduce the risk inherent in business relationships.

The latest trends in the RM research include Petison [54] researching the problems of marketing relationships between business partners from different cultures through the example of Thai suppliers and Japanese customers. Petison states that differences between individual cultures increase the number of conflicts between the parties, thus hindering the development of the business relationship. Rakshit and Eyob [55] studied the impact of RM on the quality of customer-to-customer relationships and on customer loyalty in the Ethiopian mobile telecommunications market. They state that personalized product/service plays a key role in the development of business relationships in the telecommunications sector. RM orientation and practice of Nigerian banks are examined by Olotu, Maclayton and Opara [56]. In the studied country, they find shortcomings in the practical application of the RM orientation and they urge its development. Omar and Ali [57] are trying to determine the role of brand loyalty and RM in the Islamic banking system. They emphasize that employees have a key role in strengthening the positive relationship with customers and have a significant impact on the structure and culture of the company.

1.5 Factors Affecting the Development of the Agency-Client Relationships

The agency-client life cycle theory is such a process-oriented approach, which emphasizes that cooperation and development of relationships are affected by many factors. In the process of cooperation, a number of objective and subjective factors may cause significant changes, such as the change of the customer's attitude to the relationship. Advertisers should build good relationships with

agencies so that marketing campaigns can run successfully. The failure and termination of the relationship may cause significant time, money and energy costs – creating a new campaign, searching for a new agency as well as building trust are time-consuming. Practical experience shows that in the field of advertising market establishing new market relations that can be considered stable in the future takes about two years. It is also important to note that building relationships is a dynamic process. A company during its operation builds and terminates a number of relationships [58].

It is essential for the client to be sufficiently familiar with the agency in order to build and develop the relationship. When examining the factors affecting the life cycle, it is necessary to know the purpose the business partners are working for, what kind of values they represent, and what future development they find important. Due to the factors outlined above, as well as the diversity of enterprises, factors should be diversified.

In the course of the analysis of the agency-client relationships a distinction should be made between a long-term working relationship with the agency and when the client is just in the phase of finding a new agency.

Cagley and Roberts [59], in the selection process, also consider human factor determining. They emphasize that the success of this phase of the life cycle is determined by the quality composition and compatibility of the agency´s staff. According to the authors, advertising agencies are businesses providing specialized services, and their successful operation is greatly affected by staff performance and customer satisfaction.

According to Davies and Prince [60], there exist key actions, which an agency must perform in order to successfully meet the needs of the clients. The authors emphasize the following activities as crucial: personal relationship with senior management, positive customer references, publication of successful campaigns, to facilitate the launch of a new business from the side of marketing, direct mail, and organizing business meetings.

Fam and Waller [61], in their research on agency-client relationship, distinguished four phases of the life cycle (initial phase, development phase, maintenance phase, and termination of the relationship), and they analysed the roles of a number of factors affecting certain life stages. The authors examined the factors that influence the formation of business relationships at the early stages of the relationship, while in the other three life-stages they focused on the role of factors affecting the joint work. Our own research also examines the effects of factors on relationships accordingly we found it important to present the factors in the table below.

Table 1
Description of factors influencing the selection of an agency

Factors	Description of factors
Resources of the agency	The volume of goods and services of the agency, sales strength, stability and uniqueness.
The agency´s reputation	Clients´ reviews of the agency, openness to further cooperation.
Marketing and strategic development	It includes the agency´s attitude towards the clients, as well as providing the necessary marketing tools in order to implement and achieve the predetermined project objectives.
Honesty and common goals	This factor involves the agency's ability, by which it is able to set in order to achieve the goal.
Creativity	Creative solution ideas and their implementation
Interpersonal relationships	The agency's ability to take into account customers´ needs and it creates for them the most optimal conditions of its business policy.
The quality of human resources	The quality of human resource depends on the managers working at different levels as well as on the joint work of other departments.
The agency's experience	The ability to handle problems that emerge during the implementations of the project.

Source: [61]

2 Objectives and Methodology of Research

This current study provides partial results of an ongoing research aimed at analyzing the most important factors influencing the relationship between advertising agencies and their clients in Slovakia. Our most important objective was to explore the factors that are important in the initial phase of an agency-client relationship, and to determine which factors influence the selection of the agency in the advertising industry. Our research is closely linked to the study of Fam and Waller in 2008 [61], joining their research group, which examined these issues in a very different environment, in the Far East. During the empirical research we used the authorized scales of Fam and Waller which have already been tested and successfully applied by them. Authors of the research assigned 33

items around the 8 determinant factors of the selection, and asked their respondents from Hong Kong to assess their statements in relation to the selection (1-not important, 7-very important). The novelty of our research is that we measured the factors determining the relationships between the participants of the advertising market in a different geographical, social and economic environment. Slovakia is a moderately developed country with multinational companies and multinational advertising agencies in its advertising market. As part of the international research project, in order to be able to compare the results, the same questionnaire was used to explore the importance of factors. The questionnaire was provided by one of the coordinators of the original research-Shyam Kim Fam.

Advertising agencies and their clients were involved in the research. In the questionnaire, we aimed to compare the viewpoints of the agencies and their clients in the early stages of their business relationship. In order to collect data, a database of advertising agencies operating in the markets of the country was needed. The Internet and the associations of advertising agencies of Slovakia were our sources. Then, we sent our questionnaire online to 1550 companies. The questionnaire was completed by 186 companies. Willingness to answer the questions was 12% which is appropriate. In general, respondents either reluctantly participate in the survey or are afraid that competitors obtain their data. One-fifth of respondents (186 companies) who completed the survey were the agencies and 80% of respondents were their clients. Surveyed companies first had to answer the question regarding their relationship with their agency/client, that is, at which stage of the life cycle their relationship is. After having answered this screening question, in 33 statements respondents evaluated the factors determining the initial phase of the relationship. Regarding life cycle stages, respondents were asked to evaluate our statements from 1 to 7. Analyses were undertaken using MS Excel and SPSS.

3 Results

Our research related hypothesis was based on the results of the research carried out earlier by Fam and Waller [61]. They identified 8 factors at the early stage of the agency-client relationship that affect the further development of the relationship: interpersonal relationships, creative skills, quality of account management, honesty and common goals, resources of the agency, marketing and strategic development, experience of the agency and last but not least its reputation.

Based on the above, our hypothesis is:

H: The initial phase of the advertising agency-client relationship is influenced mostly by the development of interpersonal relationships.

Since the table for 33 statements would be too large, first we continue to carry out our factor analysis.

Before carrying out the analysis, we looked at our data whether they are suitable for the analysis. All 33 statements are measured on a metric scale and our sample size exceeds 100. First, we examined the correlation between the statements. The correlation was accepted with a five per cent significance level. Principal component analysis was selected as a data reduction procedure to find out background characteristics. First, we had to decide whether our data is suitable for factor analysis. To do this, we determined the correlation matrix with SPSS program, and we analyzed the MSA value of the anti-image correlation matrix, the results of Bartlett's test, as well as the value of KMO. Homogeneity of the sample is an important condition of the implementation of factor analysis. We assume that our target groups (agencies-clients) evaluate each variable differently. Therefore, factor analysis was carried out with all of our advertising agencies and their clients. In the course of determining the correlation matrix, of the 528 correlation value, 481 (91%) was significant, and the highest correlation value is 0.77 between the statements - the existence of international relations and the importance of business growth. Based on the correlation values, we can state that the variables are suitable for factor analysis. Based on the anti-image matrix, MSA values are between 0.469 and 0.174, Bartlett´s test ($\chi^2 = 4578.511$, df = 528, p = 0.000) is significant, that is, the variables are suitable for factor analysis, as they correlate with each other. Finally, the value of KMO is 0.646, which means that the variables are moderately suitable to conduct factor analysis. We can state that, based on the indicators above, our variables are suitable for factor analysis.

When determining the number of principal components, first we analyzed our variables on the basis of the Kaiser criterion, and as a result we define 10 factors. These ten factors explain the total variance of 71.3%, which corresponds to the 60% desirable minimum value. Thus, both the Kaiser criterion and the method of variance ratio motivate us to create 10 factors. Conducting the rotation of factors, which was carried out with orthogonal rotation method with varimax procedure, we decided on the ten-factor solution. Factors, variables and means are summarized in the table below (Table 2).

Table 2

Factors influencing the early stages of business relationships

Factors	**Variables**	**Mean Agency**	**Mean Client**
Exactitude **Mean – Agency 5,12** **Mean – Client 4,89**	Business growth	5,47	5,03
	International relations	5,32	4,72
	Account management	4,55	4,87
	Cost-conscious management	4,86	4,82
	Project registration	5,41	5,00
Experience	Company size, product range	3,41	3,58

Mean – Agency 4,57 **Mean – Client 4,73**	Skills	4,64	4,65
	Support in marketing plan	4,91	4,97
	Market research capabilities	4,95	5,43
	Personnel skills	4,95	5,00
Interpersonal relationships **Mean – Agency 5,51** **Mean – Client 4,95**	Personnel compatibility	6,00	5,15
	Synergy rate	5,36	4,87
	Media at favorable rates	5,18	4,82
Ingenuity **Mean – Agency 5,00** **Mean – Client 4,93**	Creative product	5,14	5,05
	Creative strategy	4,95	5,02
	Integration of media functionality	4,91	4,73
Resources **Mean – Agency 5,35** **Mean – Client 4,85**	Creativity, research and marketing	5,55	4,97
	Sufficient resources	5,14	4,72
Survival skills **Mean – Agency 4,91** **Mean – Client 5,11**	Customer recommendations	5,64	5,20
	Flexible pricing	4,77	5,08
	Reasons for the termination of a relationship	4,32	5,05
Customer care **Mean – Agency 5,00** **Mean – Client 4,82**	Marketing and communication services	5,18	4,97
	Top management participation	4,95	4,66
	Quality of account management	4,86	4,83
Stability **Mean – Agency 4,95** **Mean – Client 4,67**	Received awards	4,82	4,53
	Workforce stability	5,27	4,72
	Support in long-term business strategies	4,77	4,77
Willingness to transfer information **Mean – Agency 5,15** **Mean – Client 5,02**	Draws attention to bad decisions	5,14	5,08
	Information on global market trends	4,82	5,12
	Agreement on objectives	5,50	4,87
Stability preservation **Mean – Agency 5,13** **Mean – Client 5,04**	Appropriate handling of account data	5,45	4,90
	Development of strategic skills	4,59	5,10
	Creative philosophy	5,36	5,13

Source: own

When analyzing the factors, those rates were taken into account that exceed the value of 0.3 due to the sample size. The following ten factors have been emphasized: exactitude, experience, interpersonal relationships, ingenuity, survival skills, customer care, resources, willingness to transfer information, and stability preservation. Based on the factor analysis, we can state that the early stage of a relationship is determined by exactitude (10%), experience, interpersonal relationships, ingenuity, resources and survival skills (7-7%), customer care, stability, willingness to transfer information (6-6%) and stability preservation (5%).

Based on the table above, except for two factors (experience and survivability), agencies consider the role of the outlined factors in the development of the relationship more important than their clients do. We can see outstanding differences in the assessment of the existence of resources. While agencies, in the development of business relationships, consider their creativity, research activities, marketing work and their available resources very important, their clients find this group of factors less important. There is also a big difference in the perception of the role of interpersonal relationships in the development of business relationships. The development of interpersonal relationships, their existence and retention are considered substantially more important by the agencies. Our hypothesis that interpersonal relationships play a crucial role in the development of business relationships in the advertising market, has not been proved since the participants of both the demand and the supply side consider another criterion the most determining at this stage. While the agencies consider interpersonal relationships, especially the degree of compatibility and synergy between them as the most influential factor, clients find the agency's ability to survive that is, for example, the recommendations of former clients more important. Our hypothesis, which states that *the initial phase of the advertising agency-client relationship is influenced mostly by the development of interpersonal relationships*, is disapproved.

The importance of interpersonal relationships at the initial stage of cooperation is considered differently by the group of agency and the group of clients.

Transfer of information is also very important in the future of a good relationship. Regarding development, the majority of respondents considered the listed factors rather important than less important.

In the next phase of the study we wanted to find out whether there is a difference between customer and agency reviews regarding the assessment of the factors influencing the initial phase of cooperation. Analysis of variance was used to compare the two groups. This method is for testing differences between the means of the two groups. Having examined the conditions of analysis of variance, it is true that the values of dependent variables are in the interval. For homogeneity of variances Levene's test was used. After running the test program, our findings showed that homogeneity of variances in the case of two factors – exactitude and survival skills - is not fulfilled. Thus, the ANOVA table 3 contains 8 factors.

Table 3

Assessment of factors influencing the initial phase of cooperation from the aspect of customers and agencies

Factors	**F**	**Sign.**
Experience	4,101	0,044
Interpersonal relationships	18,577	0,000
Ingenuity	1,028	0,312

Resources	2,566	0,110
Customer care	0,789	0,375
Stability	0,702	0,403
Willingness to transfer Information	0,001	0,974
Stability preservation	0,485	0,487

Source: own

Significance levels indicate that agency and client group consider experience and interpersonal relationships in the initial phase of cooperation important to varying degrees. Ingenuity, resources, customer care, stability, willingness to transfer information, and the importance of preserving stability are evaluated the same way by both groups. Accordingly, the agency and the group of clients consider the importance of interpersonal relationships in the initial phase of cooperation differently.

Conclusions and Managerial Implications

This study aimed to analyze the selection of advertising agencies from the perspective of agencies and customers. Our findings disprove the previous research results, according to which interpersonal relations and human resource play the main role in the development of business relationships in the advertising market. Our surveyed advertising agencies agree that human resource is the most important factor (creativity, research and marketing work are necessary requirements). It is, therefore, important that at the initial stage of developing business relationships agencies focused on demonstrating their creative skills, paid due attention to clients´ problems, which may be a condition for preserving the relationship. Furthermore, according to them, advice and open communication are expected by the clients. On the contrary, on the basis of the reviews of the Slovak advertising market´s customer base we can state that clients´ recommendations, flexible remuneration, as well as the knowledge of the previously lost clientele and reasons for terminating the relationship are more determining factors in the selection of an agency. So, we can state that if we accept that the advertising market is also customer-centric, agencies should pay much more attention to the publication of their clients´ recommendations when advertising their work, either on the website of the company or by words of mouth. Distrust in business partners, that is typical of the Slovak corporate sector, may be moderated by previous recommendations. In our opinion competition may prompt new solutions and the development of even stronger trust in relationships in the advertising market.

References

[1] L., Mura, M., Lincényi: *Základy marketingu.* (Principles of Marketing) Brno: Tribun EU, 2015

[2] J. E., McCarthy: *Basic Marketing. A Managerial Approach.* Homewood, III., R. D. Irvin, 1960

[3] C., Grönroos: From marketing mix to relationship marketing – towards a paradigm shift in marketing. *Management Decision,* Vol. 32 (2), 1994, pp. 4-20

[4] K., Grayon, T., Ambler: The dark side of long-term relationships in marketing services. *Journal of Marketing Research*, Vol. 36 (1), 1999, pp. 132-141

[5] E., Gummesson: *Total Relationship Marketing: Rethinking Marketing Management from 4Ps to 30Rs.* Oxford: Butterworth Heinemann, 1999

[6] T. L., Baker, P. M., Simpson, J. A., Siguaw: The impact of suppliers' perceptions of reseller market orientation on key relationship constructs. *Journal of the Academy of Marketing Science*, Vol. 27 (1), 1999, pp. 50-57

[7] K. S., Fam, D. S., Waller: Factors in winning accounts: The views of New Zealand agency account directors. *Journal of Advertising Research*, Vol. 39 (3), 1999, pp. 21-32

[8] L. L., Berry, A., Parasuraman: *Marketing services—Competing through quality.* New York: Free Press, 1991

[9] N., Alexander, M., Colgate: Retail financial services: transaction to relationshipmarketing. *European Journal of marketing*, Vol. 34 (8), 2000, pp. 938-953

[10] L. L., Berry: Relationship marketing. In: Berry, L. L., Shostack, G. L. & Upah, G. D. (Eds.) *Emerging Perspectives of Service Marketing.* Chicago: American Marketing Association, 1983, pp. 25-38

[11] A., Sharma, J., Sheth: Relationship marketing: An agenda for inquiry. *Industrial Marketing Management*, Vol. 26(2), 1997, pp. 87-89

[12] J. N., Sheth: *Relationship Marketing: Paradigm Shift or Shaft. In J. N. Sheth and A. Parvatiyar (Eds.), Handbook of Relationship Marketing*, Thousand Oaks, CA: Sage Publications, 2000, pp. 609-620

[13] T. S., Kuhn: *The Structure of Scientific Revolutions*. Second Edition, Enlarged, The University of Chicago Press, Chicago, 1962

[14] M. J., Harker, J., Egan: The Past, Present and Future of Relationship Marketing, *Journal of Marketing Management*, Vol. 22, 2006, pp. 215-242

[15] Z., Francová, M., Oreský: *Vzťahový marketing: od teoretických východísk po uplatnenie v praxi*. (Relationship marketing: from theory to practice), Bratislava: Vydavateľstvo EKONÓM, 2010, p. 313

[16] G. S., Day: Managing market relationships. *Journal of the Academy of Marketing Science*, Vol. 28 (1), 2000, pp. 24-30

[17] L. A., Crosby, K. R., Evans, D., Cowles: Relationship quality in services selling: An interpersonal influence perspective. *Journal of Marketing*, Vol. 54, 1990, pp. 68-81

[18] R., Priluck: Relationship marketing can mitigate product and service failures. *Journal of Services Marketing*, Vol. 17(1), 2003, pp. 37-52

[19] P., Kotler, K. L., Keller: *Marketing Management.* Pearson Prentice Hall, 2006

[20] N., Sharma, P. G., Patterson: Switching costs, alternative attractiveness and experience as moderators of relationship commitment in professional, consumer services. *International Journal of Service Industry Management,* Vol. 11 (5), 2000, pp. 470-490

[21] P. S., Gounaris: Trust and commitment influences on customer retention: insights from business-to-business services. *Journal of Business research,* Vol. 21 (4), 2005, pp. 126-140

[22] B., Vatanasombut, M., Lgbaria, A., Stylianou, W., Rodgers: Information System Continuance Intention of Web Based Applications Customers: The Case of Online Banking. *Information and Management*, Vol. 45 (7), 2008, pp. 419- 428

[23] Ch., Sichtmann: An analysis of antecedents and consequences of trust in a corporate brand. *European Journal of Marketing*, Vol. 41 (9/10), 2007, pp. 999-1015

[24] J. C., Anderson, J. A., Narus: A model of distributor firm and manufacturer firm working partnerships. *Journal of Marketing*, 54 (1), 1990, pp. 42-58

[25] L., Lages, J., Abrantes, C., Lages: The STRATADAPT scale: a measure of marketing strategy adaptation to international business markets. *International Marketing Review,* Vol. 25 (5), 2008, pp. 584-600

[26] B., Ashnai, M., Smirnova, S., Kouchtch, Q., Yu, B. R., Barnes, P., Naudé: Assessing Relationship Quality in Four Business-to-Business Markets, *Marketing Intelligence & Planning,* Vol. 27 (1), 2009, pp. 86-102

[27] P. J., Danaher, A., Bonfrer, S., Dhar: The Effect of Competitive Advertising Interference on Sales for Packaged Goods. *Journal of Marketing Research,* Vol. 45, 2008, pp. 211-225

[28] M. T., Aziz, N. A. M., Noor: Evaluating the effect of cost related factors on relationship quality: An investigation of retailer- supplier relationship in Bangladesh, *International Journal of Retail & Distribution Management*, Vol. 41 (7), 2013, pp. 545-558

[29] T., Gao, M. J., Sirgy, M. M., Bird: Reducing buyer decision-making uncertainty in organizational purchasing: can supplier trust, commitment, and dependence help? *Journal of Business Research*, Vol. 58 (4), 2005, pp. 397-405

[30] L. Y. M., Sin, A., Tse, F., Yim: CRM: conceptualization and scale development. *European Journal of Marketing*, Vol. 39 (11-12), 2005, pp. 1264-1290

[31] C., Perry, A., Cavaye, L., Coote: Technical and social bonds within business-to-business relationships. *Journal of Business & Industrial Marketing,* Vol. 17, 2002, pp. 75-88

[32] L. V., Coote, E. J., Forrest, T. W., Tam: An investigation into commitment in non-Western industrial marketing relationships. *Industrial Marketing Management*, Vol. 32 (7), 2003, pp. 595-604

[33] U. R., Orth, M. T., Green: Consumer loyalty to family versus non-family business:The roles of store image, trust and satisfaction. *Journal of Retailing and Consumer Services,* Vol. 16 (4), 2009, pp. 248-259

[34] V., Kumar, D., Shah: Building and sustaining profitable customer loyalty for the 21st century. *Journal of Retailing*, Vol. 80, 2004, pp. 317-330

[35] R., Lacey, R. M., Morgan: Customer advocacy and the impact of B2B loyalty programs. *Journal of Business and Industrial Marketing*, Vol. 24 (1), 2009, pp. 3-13

[36] G., Shi, H., Bu, Y., Ping, M. T., Liu, Y., Wang: Customer relationship investment and relationship strength: evidence from insurance industry in China, *Journal of Services Marketing*, Vol. 30 (2), 2016, pp. 201-211

[37] R., Kotorov: Ubiquitous organization: Organizational design for e-CRM. *Business Process Management Journal*, Vol. 8 (3), 2002, pp. 218-232

[38] L. L., Kelley, D., Gilbert, R., Mannicom: How e-CRM can enhance customer loyalty. *Marketing Intelligence & Planning*, Vol. 21 (4), 2003, pp. 239-248

[39] L., Da-wei: Study on the factors of customers' loyalty in e-business world. In *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, China, 2007, pp. 34-74

[40] S. C., Sivakumar: A critical survey of protocols proposed by the IETF as enablers for customer interaction in an electronic customer relationship management system – Part I – eCRM Metrics. *IEEE Canadian Review*, 2002, pp. 5-9

[41] A., Payne, P., Frow: A strategic framework for customer relationship management. *Journal of Marketing*, 2005, Vol. 69 (4), pp. 167-176

[42] R., O'Callaghan, P. J., Kaufmann B. R., Konsynski: Adoption correlates and share effects of electronic data interchange systems in marketing channels. *Journal of Marketing*, Vol. 56 (2), 1992, pp. 45-56

[43] K. A., Patterson, C. M., Grimm, T. M., Corsi: Adopting new technologies for supply chain management. *Transportation Research Part E, Logistics and Transportation Review*, Vol. 39, 2003, pp. 95-121

[44] L. Y. M., Sin, A., Tse, H., Chan, V. C. S., Heung, F., Yim: The effects of relationship marketing orientation on business performance in the hotel industry. *Journal of Hospitality & Tourism Research*, Vol. 30 (4), 2006, pp. 407-426

[45] R. V., DaSilva, S. F., SyedAlwi: Cognitive, affective attributes and cognitive behavioural responses in retail corporate branding. *Journal of Product and Brand Management*, Vol. 15 (5), 2006, pp. 293-305

[46] R. L., Oliver: Whence consumer loyalty. *Journal of Marketing*, Vol. 63 (4), 1999, pp. 33-44

[47] R. W., Palmatier, S., Gopalakrishna, M. B., Houston: Returns on business-to-business relationship marketing investments: strategies for leveraging profits. *Marketing Science*, Vol. 25 (5), 2006, pp. 477-493

[48] R. W., Palmatier, L. K. Scheer, J. B. E. M. Steenkamp: Customer Loyalty to Whom? Managing the Benefits and Risks of Salesperson-Owned Loyalty. *Journal of Marketing Research,* Vol. 44, 2007, pp. 185-199

[49] C., Homburg, J., Wieseke, W. D., Hoyer: Social identity and service-profit chain. *Journal of Marketing*, Vol. 73 (2), 2009, pp. 38-54

[50] M., Huang: The influence of selling behaviors on customer relationships in financial services. *International Journal of Service Industry Management*, Vol. 19 (4), 2008, pp. 458-473

[51] R., Bennett, S., Rundle-Thiele: Customer satisfaction should not be the only goal. *Journal of Services Marketing*, Vol. 18 (7), 2004, pp. 514-523

[52] R. M., Morgan, S. D., Hunt: The commitment-trust theory of relationship marketing. *Journal of Marketing*, Vol. 58, 1994, pp. 20-38

[53] K., Nykanen, K., Moller, T., Järvensivu: Towards a more social perspective of network management: An action research study on trust and commitment. In *Proceedings of the 25th IMP Conference*, Marseille, France, 2009

[54] P., Petison: Cross cultural relationship marketing in the Thai context: The Japanese buyer's perspective. *International Journal of Trade . Economics and Finance*, Vol. 1 (1), 2010, pp. 17-23

[55] N., Rakshit, K., Eyob: Relationship marketing and customer loyalty: The Ethiopian mobile communications perspective. *International Journal of Mobile Marketing*, Vol. 5 (1), 2010, pp. 113-124

[56] A. O., Olotu, D. W., Maclayton, B. C., Opara: An empirical study of relationship marketing orientation and bank performance. *Research Journal of International Studies*, Vol. 16, 2010, pp. 47-57

[57] M. W., Omar, M. N. M., Ali: Brand loyalty and relationship marketing in Islamic banking system. *Canadian Social Science*, Vol. 6 (1), 2010, pp. 25-32

[58] D. S., Waller: Developing an account management lifecycle for advertising agency–client relationships. *Marketing Intelligence & Planning,* Vol. 22 (1), 2004, pp. 95-112

[59] J., Cagley, C. R., Roberts: A comparison of advertising agency selection factors. *Strategic Managment Journal*, Vol. 24, (2), 2004, pp. 39-44

[60] M., Davies, M., Prince: Dynamics of Trust Between Clients and Their Advertising Agencies: Advances in Performance Theory. *London: Academy of Marketing Science*, 2005, Vol. 8 (1), pp. 12-29

[61] K., Fam, D., Waller: Agency–Client Relationship Factors Across Life-Cycle Stages. *Journal of Relationship marketing*, Vol. 7 (2), 2008, pp. 217-236

A Contribution to Software Development Quality Management

Vidan Marković, Zora Konjović

Faculty of Organizational Sciences, Chair of Information Systems, University of Belgrade, 154 Jove Ilica, 11000 Belgrade, Serbia
University Singidunum, 32 Danielova, 11000 Belgrade, Serbia
vidan.markovic@fon.bg.ac.rs, zkonjovic@singidunum.ac.rs

Abstract: Disruptive digital technologies have increased the potential for international businesses to access clients' data in traditionally closed local markets (e.g. insurance, banking, etc.) without the need for a physical presence. In order to stay competitive in the long run, local companies have to invest more in both technology and sourcing strategies. In this paper, we propose an enhancement to the software development suppliers management process based on a continuous services comparison that can lead to proactive improvements in suppliers' quality of services. This paper explains this process and demonstrates a company case study experiment.

Keywords: Software Quality; Software Processes; Outsourcing

1 Introduction

The ability to react fast to a recognized need for change of the functionality or performance of an existing software service, or to an emerging business need for a new software service running in production, is becoming increasingly important [1]. However, obtaining and keeping the right resources to do this job is increasingly difficult and complex. Thus, outsourcing in software development has become inevitable in order to keep pace with these needs [2]. This means that the in-house software development process must adapt to this new reality as well. Handling both new software projects development and the maintenance of existing software services through a mix of internal and external resources is not a trivial task. Companies need to update/improve their software development strategies, and practice SW development results interpretations from different perspectives: from an immediate goal achievement to wider strategic alignments with business needs.

What really matters at the end of any software development project is the perceived value of a production and the quality of IT services. These basic

objectives must be the top priority during the whole software development supplier engagement life cycle: from the supplier selection process, to on-going operations, to the supplier's contract closure.

The internet of things, social networking, big data analysis, cloud technologies, and high mobility all require that user interfaces be available on various platforms and in various environments at any time. These developments have only increased the need for better management of simultaneous projects' developments, technical implementations, and operations' teams [3]. Agility, internal and external collaboration (with users, customers, partners, competitors, regulatory bodies, etc.), and the need for frequent deployments have also increased demand for better security procedures, policies, and tools.

In order to keep pace with these exponentially increased expectations of businesses, CIOs have to rethink and continuously improve IT human resources sourcing models and try to find the best fit solution for their particular companies and eco-system needs. This means creating a desirable balance between functionality, quality, costs, and risk targets for the company [4].

Even though the number of standardized software solutions for different business industries and the business processes within them are growing [5] (many are already available as SaaS Cloud solutions [6]), companies still strive to get differentiation in the market and gain competitive advantage from the specifics of their own solutions.

In this paper, we discuss the question of how to measure and manage the quality of tailored business software solutions when software is increasingly being developed outside of company control and the gap between knowledge inside the company and the knowledge of the code is growing. Finally, how do you mitigate that kind of risk for the business?

In order to answer this, we propose utilization of the Six Step Service Improvement (3SI) method [7] in managing quality, costs, risks, and general relationships with programming outsourcers. This method enforces continuous communication with outsourcers, which is a particularly important factor that influences the quality of software services.

The 3SI method can be used in supporting transparent, performance based relationships with suppliers. However, in order to gain the most benefit from using this method, it is very important that the IT organization is mature and capable enough to professionally plan, execute, and control products and services procurement in its best interest (based on predefined procedures, polices, and expectations fulfillment criteria).

We verified this method with a case experiment on a selected instance of the use cases' class. This case company was characterized by a relatively stable business environment, by the internal processes, organization, and culture of an experienced team, and by a relatively stable system architecture (company was at

CMMI level 4). However, this method could be applied in the various implementation scenarios (i.e. different business domains/eco systems, different companies' CMMI levels, different programming languages, etc.).

This paper begins in this introduction section by defining the problem and providing a brief summary of our work, then continues through an overview of related work and the hypothesis of the potential solution. After that, it gives a detailed explanation of the proposed solution based on the utilization of the 3SI methodology. It continues by describing the results tested and verified in the case study experiment, along with a description of the benefits of the proposed method utilization. Finally, the paper finishes with a conclusion, further work proposal, and the references list.

2 Related Work and Hypothesis

Managing the requirements in iterations and building solutions based on the integration of developed and tested components and packages represent, the core of modern software development processes [8, 9]. The complexity of software solutions is growing because of an increased need to dynamically exchange information with a high number of open systems and databases.

The complexity, is also reflected in a need for higher operating capacity, performances, and highly efficient development tools and components to cover an increasingly larger area of the application domain by integrated systems. Often without proper methodological management of all possibilities and risks of new technologies.

The pressure to shorten delivery times (agile solutions to software projects) leads to a quick release of a valid version, but can lead to problems in the future maintenance if the critical knowledge about the current version of system is not systematically storied in suitable, understandable and precise form.

In order to manage the increased expectations for software services it is important to manage the associated costs and risks (that could be again mapped to the costs). COCOMO II, and other similar methods, can give a good approximation of the software cost estimation [10, 11, 12, 13, 14].

The main issue with using these methods in today's business environments is the fact that business needs fast answers that ideally could be automated and ran regularly for decision making support. Typical scientific software economic measurements and costs estimate tools (e.g. KSLOC (thousands of Source Line of Code), function points, etc.) unfortunately, do not get enough attention from business to be used in a real-time environment. Another issue with the traditional scientific approach is that the data calculated and analysed in most scenarios *do not consider the specifics of the organization business and its eco-system*

dynamics and because of that are conceived as micromanagement tools rather than strategic decision making support tools.

That is the reason why, in business practice, most estimations are based on a *rule of thumb* principle. These estimates are more accurate when applied on similar projects (i.e. the same architecture, same platform to be developed and to run on, similar teams, etc.). Previous experience and lessons learned play an important role in these calculations.

However, with an open enterprise architecture, the increasing size of externally programmed code, and without sufficient and sustainable internal knowledge, setting a reference point for the quality, costs, and risk assessments of new software scheduled for deployment becomes extremely cumbersome.

There are principles like the Quality Improvement Paradigm (QIP), that define software discipline as evolutionary and experimental [15, 16]. This means that there is very little repetition in software development, which using statistical control in software quality control, like that used in manufacturing sciences, makes it extremely difficult and dubious [11]. The developers of QIP take a different approach than, for example, the authors of the CMMI and a number of other models that are based on the idea of statistical control of processes [16].

However, meeting the quality needs of software services includes the principles written by Deming [17, 18] and found in Total Quality Management (TQM) practice. Regardless of the particular flavour of TQM implemented, process definition, control, and improvement are always included as core TQM principles [19]. The main idea behind process control is that organizations are sets of interlinked processes, and improvement of these processes is the foundation of performance improvement [20, 21].

The oldest model that can be seen as an improvement action life-cycle model is the PDCA (Plan-Do-Check-Act) model developed by Shewhart and Deming [18]. It was originally devised for improving quality in manufacturing, and has its foundation in statistical quality control, i.e. controlling quality by applying metrics to the process.

All the basic principles of quality improvements still exist, but they now must work on higher dynamics and higher complexity. The time to market criteria has a higher weight coefficient within dynamic business environments (e.g. fast changes in products portfolio's, new organizations and processes due to mergers and acquisitions, economic crises, regulation requirements, etc.).

Squale quality models are concentrated on visualization of metrics where distribution maps, tree views, tree rings, etc. help in better understanding the quality of software, specifically by adding *practice* as an intermediate level between metrics and criteria as defined by ISO 9126 (that promotes a three-level model of quality: factors, criteria, and metrics) to support improvements actions [22]. However, even though with the Squale the visualization helps in

understanding the software quality issues from different a perspective (i.e. from software development level to CIO level), it still requires significant preparation and high-level of technical expertise in understanding the potential quality issues because of its strong bottom-up approach in software quality assessment.

We believe that in operational utilization of software quality improvement technics, where decision about improvements have to be made quickly (due to business ecosystems factors), it is critical to have a balanced mix of internal complexity (white box) consideration and external results, visible to end users (black box). It means that bottom–up and top-down approaches need to be in line with specifics of the business and software development environment.

Emerging DevOps methodology integrates development and operation activities in order to frequently generate new software deployments (i.e. in matter of minutes in some cases). New tools are arising in the market to support continuous deployment with high-level of automation.

Although all the above methods and software processes emphasize the need for continuous quality improvements, there is no clear and systematic approach on how to successfully manage and achieve continuous quality improvements in continuously changing business and technical environments.

We limit our analysis for the domain of business software solutions in services oriented industries such as banking, insurance, education, government, tourism, etc., where their core processes are heavily supported and influenced by various software utilization (made in-house, bought from third parties, and mixed solutions).

Hypothesis definition: There is *a holistic solution* that supports building higher quality software (for the business domain defined above, within specific industry, for the specific business function: i.e. claims management software for claims management department in an insurance company, or child care management software for social services department in a municipal government) by means of continuous improvements based on regular/periodical (with a particular sense of the business's *pulse*) utilization of the given methodology on defined sets of empirical, real-time data generated during software development and operations processes.

3 The Solution Proposal

"An incident is an unplanned disruption or degradation of service. A problem is a cause of one or more incidents. Quite often, in operations, these two terms are used interchangeably" [23, 24].

Minor outstanding incidents in some parts of the software product will be naturally discovered in production system utilization because with the higher complexity of software systems, there is a higher probability that some bugs will pass through all predefined levels of testing and quality assurance. That is the first reason why continuous improvements are necessary.

The second reason for software changes is a result of changes in the operating environment (i.e. infrastructure changes, operating system changes, other interconnected systems' changes, interfaces' changes, etc.) that require software adaptation to the new production environment.

Performance-related reasons also influence software changes. There are the "*Would like to have* type" of end-users' requests (MoSCoW rule) that are often purposely left as non-critical and dealt with in future software releases.

However, the most significant change requests come from a business's new/changed functional needs for a software service. More than 15% of software defects are related to requirement's errors [10]. Errors that were not detected in earlier phases of software development contribute to a higher cost in fixing defects later on [8, 14].

If, besides the above reasons for software services changes, a software service suffers unexpected degradation of quality (e.g. more bugs, more production fixes, non-compliant SLA, etc.), higher development and/or maintenance costs and risks, and if similar incidents repeat in the same or other parts of the system, then this would normally need to be escalated as an issue [24, 25]. Moreover, if the software program code has been signed by the same supplier of the programming services, and there is a recognizable pattern (rather than just a normal variation of errors within predefined statistical control boundaries [17, 23, 26]), then the company has a *problem* that needs to be resolved.

The normal engineering tendency is to technically rationalize increasing numbers of incident occurrences as (for example): too many changes on initial requirements, no ability or stability to define firm scope, poor business analysis, not properly done design (e.g. no UML diagrams, just a simple user-story or some kind of specification with a fractions of pseudo-code), no standards, etc. However, despite this rationalizing tendency, there are reasons to believe that in many cases the root of the problem was in communications' procedures and policies pitfalls.

Communication issues cannot only enlarge a small incident, but if dealt with properly, can also effectively solve a big one. Thus, treating communication as a main tool for supplier management will help in solving problems with outsourcers, especially in the area of supplier's quality of service expectations.

Agile methodologies (e.g. SCRUM, Dynamic Systems Development Method (DSDM), Extreme Programming (XP), etc.) emphasise utilization of suitable prototypes [27, 28] and effective mapping of user stories [27, 29] to improve

communications of centralized or distributed development teams and users, and to lower development risks.

In most scenarios, in order to solve a complex problem it is a good idea to decompose the whole problem into smaller, better manageable, and understandable parts [9, 29, 30]. This means both simplifying and diversifying the views of the existing problem. We should limit our analysis of programming services outsourcers as manufacturers of these parts.

We also believe that a *holistic view of the problem* needs to be introduced in order to change the predominant focus on the technical side of each single instance of the problem class (that are normally hard to track [31, 32, 33]) to the more statistical and processes sides of the problem as a whole (i.e. a view of the problem class conducted in parallel with a view of a single problem instance).

Thus, a holistic view of the project set-up and maintenance process set-up with suppliers can be a prerequisite for further analysis. The main goal would be to build mutual trust and improve the quality of overall services. This also means moving outstanding issues (errors/bugs) back to predefined and mutually agreed limits of the statistical control.

In addition, we propose that with continuous cooperation on improving predefined KPIs (Key Performance Indexes), the agreed limits of statistical control could also become a target for improvements. This plan could become a common part of contracts, with the idea of rewarding high quality in the interest of both parties (i.e. introduction of incentive types of contracts rather than time and material or fixed term). Thus, software problems would be first transferred into the relationship and provider management domains for a solution, and then, once solved in the soft (people) problem area, transferred back to the hard (technology) problem area – the program code.

This holistic approach does not substitute inner software development process, regardless of the type: agile (SCRUM, DSDM, etc.) or procedural (e.g. Rational Unified Process) for whatever reasons they were chosen for the particular software development. It is a better tool to manage new developments based on the opportunity designed in a systematic, methodological way to learn from production behaviour of software and humans that use them in the particular environment. It is more like an attempt to prescribe a particular patient continually improved medicines based on the data collected on his reactions over time on different medicines (and their ingredients) for the particular disease.

We have selected the 6 Step Service Improvement methodology [7] as a method that could support systematic and transparent improvements in suppliers' services.

The main characteristics of the method are:

- The result of method utilization is an improvement in the overall quality of the software service.

- It does not pre-estimate services without measuring their outcomes first. This means that nobody is the best supplier by definition. Measurements taken during selected real time production period cycles (e.g.. month, quarter, half year, and year) in the given environment will give better answers than forecasts based only on past experiences in different environments.
- The method uses a ranking principle for services groupings in order to provide meaningful comparisons among similar type of services within the same rank (apples to apples).
- The method uses LSP (Logical Scoring of Preferences) for comparison purposes among predefined elementary criteria.
- The method is conducted in improvement cycles until it makes sense to continue to another cycle (cost/benefit sense) to avoid "gold plating" scenarios. Kaizen practice is good thing to do, but it has a cost side to be addressed as well [31].
- Learning about services is continuous, and the view of the services is holistic. This means that the method supports a continuous improvements paradigm and embraces uniqueness of each service, as well as their shared characteristics. Each service is described by the dimensions of soft (people interactions) and hard (technical, i.e. code that works) elements. These elements are interconnected in multidimensional services' cubes. Recognition, based on measurements and comparison, of what makes some elements of one cube better than another helps in improving the other service and vice-versa.

We propose the use of the method for multidimensional comparison of preferences in spiral cycles in repeatable time cycles that reflect the particular *business pulse* (dynamics of the business changes). In some cases where, for example, the business pulse is at the *elephant* level (25-35 bpm resting), the time cycle periods will be longer (i.e. quarterly or semi-annually); on the contrary, for a business with the *mouse* pulse level (450-750 bpm resting), weekly cycle periods might be the right choice.

Seasonality effects in each business would need to be taken in account; thus some *arrhythmic* cycles could be desirable as well (i.e. monthly cycles normally, and during the summer only one quarterly cycle).

4 Case Study Experiment and Results

The 3SI method starts with identification of the SW services that can be classified within the same rank. As mentioned above, the quality of software services is strongly influenced by service providers' characteristics. In order to compare

service providers' characteristics we first rank them based on the *flavour* of their software services domains (i.e. programming language, system architecture environment, team structure, etc.), and then drill into further hierarchical decompositions until we reach elementary criteria [34]. The formula to calculate the estimates of each defined criteria [34, 35] is given below (1):

$$\mathrm{E} = \left(\sum_{i=1}^{k} \omega_i \; e_i^r\right)^{1/r}, 0 \leq \omega \leq 1, \; \sum_{i=1}^{k} \omega = 1,$$

$$e \in < 0; 1 >, \; \mathrm{E} \in < 0; 1 >, \; k \geq 2 \tag{1}$$

where the coefficient "ω" represents the weight coefficient associated with the comparative importance of each estimated elementary preference belonging to the same hierarchical group preference, and the "e" represents an elementary preference estimate. The "k" represents the number of features in the aggregation blocks. The "r" represents the correlation function to be applied on the specific level. The values of "r" are defined on the basis of the expectation of the combined influence of the estimated preference at the group level (e.g., synergy effects). The values for "r" vary from full conjunction (C, r = - ∞) to full disjunction (D, r = + ∞). The arithmetic mean (AM) is given at r = 1. More details about the mathematical aspects of the LSP method can be found in [36]. The strength of LSP resides in the power to model different logical relationships [34]:

- Simultaneity: when it is perceived that two or more input preferences must be present simultaneously,
- Replaceability: when it is perceived that two or more attributes can be replaced (there exist alternatives, i.e., the low quality of an input preference can always be compensated by the high quality of some other input),
- Neutrality: when it is perceived that two or more input preferences can be grouped independently (neither conjunctive nor disjunctive relationship),
- Symmetric relationships: when it is perceived that two or more input preferences affect evaluation in the same logical way (though possibly with different weights),
- Asymmetric relationships: when mandatory attributes are combined with desirable or optional ones.

The service grouping, as the first step of identification, was done based on the identified service class's group attributes [7]:

- Technology group (TDi) - represents the technical attributes that better describe the influence of applied technology tools on service development and operations.
- Complexity group (Ci) - represents the observed level of complexity in creating a solution. More tiers in the solution implementation in most cases represent more complexity in operating that service.

- Development process group (DPi) - represents the possibility to lever influence on the service by an applied development process. Some development processes could create a very stable service, but have a problem with the low level of flexibility to change.
- Development team group (DTi) - team experiences, skills, and cohesion, and in-house and outsourced options affect the ability for quality maintenance on a specific service.
- Business support domain group (BDi) - related to the end user profile, the number, location, and type of application being used (e.g. OLTP, reports, etc.).

Based on the above definitions of group attributes, each instance of service class S_i from the catalogue was assigned values as following (2):

$$S_i = (TD_i, C_i, DP_i, DT_i, BD_i),\ i \geq 0 \quad (2)$$

We analysed the use case in which there are more teams (in-house and outsourced teams) working on the same software and hardware platform, but on different projects. Thus, these projects are *environmentally* similar, but size and complexity vary depending on the needs of the end users. The end user stakeholders are employed with the same company. Internal variations within the internal teams were minimal.

The observed company uses services from four Java outsourcing providers. The decision to use four outsourcing companies for the same area of expertise came from the need to better manage the risks of outsourcing. The company has gradually moved its strategy towards Java outsourcing from pure in-house development. The reason for that strategic switch was an increased need for faster and better solutions from one side, and the limited resources in highly competitive Java programmers market from the other (unfavourable supply/demand ratios).

Managing different outsourcing companies in supporting the same or similar business area is not trivial. Differences could be in (but not limited to):

- internal development processes specifics (e.g. agile, procedural, mixture),
- design and coding standards,
- IDE (Integrated Development Environments) tools utilised,
- Version Control standards,
- project management standards,
- culture.

The company has a defined set of internal coding standards to be applied with all outsourcers, and it has also adopted standards for version control and reporting on the work assigned progress. The price of each man-day for the same type of work

(Java programming) has been negotiated to the same level for all service providers. These decisions have made it possible to concentrate on more objective KPI measurements for each code supplier.

To have a better understanding of the company's working environment, some historic facts need to be addressed as well. The company's IT department has gradually moved from pure insourcing to outsourcing of its Java programming activities (Figure 1). The main reason was an increased need for Java programming jobs without a favourable supply of Java programmers in the local market for prospective employers. On average, the retention rate was less than 3 years.

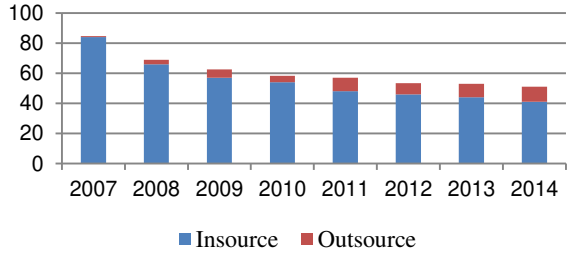


Figure 1
An example of the historical trend on insourcing vs. outsourcing FTE consumptions

However, the main issue with selecting the right balance on insourcing vs. outsourcing strategy was the tactics for managing the quality of outsourcing activities. If there are more code suppliers, some kind of multi-dimensional comparison needs to be applied that takes into account the following criteria:

- static criteria (project/activity quality of scope delivery, SPI, CPI, etc.),
- dynamic criteria (service life cycle experiences with outsourcers) including:
  - hard (number of bugs reported in period, architecture issues reported in period, costs in period, etc.),
  - soft (team relationships) criteria.

In some cases cutting off one code supplier is not feasible for different reasons (e.g. legal, no good replacements on the market, etc.). However, it is possible to learn from good experiences with different outsourcers and different services. Ideally, new knowledge on how to improve certain service parameters should be passed to that service in real time. Services should be improved continuously, ideally without any time breaks.

The suppliers reported the consumption of their Java development resources continuously for a period of one year as shown below [Table 1].

Table 1

SW suppliers efforts (in man-days) invested during observed year

Supplier	Effort (m-d)
A	458
B	724
C	307
D	263

Our primary goal was to compare the quality of services within the same service class. C_i attribute values could have significant variation depending on the type of users' requirements (e.g. functional scope complexities, non-functional technical complexities, etc.), but the other environment dependent parameter variations could be kept stable for the observed time period for the preferences calculations (i.e. $\pm$ 5%, or within the predefined result threshold).

Thus, these constants are calculated (3, 5, 7, 9, and 11):

$TD_i = (td_1, td_2, td_3)$, where $td_1 \epsilon \{2T, 3T, 4T\}$,

$td_2 \epsilon \{WO, WP, DC\}$, and $td_3 \epsilon \{J, VB, C, D\}$ (3)

Here, 2T stands for Two-Tier, 3T for Three-Tier, and 4T for Four-Tier application architecture; WO for Web/Open Source, WP for Web/Proprietary, and DC for Client (Fat Client) presentation layer; J for Java, VB for Visual Basic, C for C++, and D for DotNet programming language.

In this use case the values of TD_i domain are kept as: three-tier, Open Web platform, and Java programing language (expressions 3 and 4).

$TD_i = (3T, WO, J)$ (4)

$DP_i = (dp)$, where $dp \epsilon \{S, R, A, H\}$ (5)

Here, S stands for SSA (Structural System Analysis), R for RUP, A for Agility, and H for Hybrid.

In this use case the value of DP_i domain is hybrid 5 and 6).

$DP_i = (H)$ (6)

$DT_i = (dt)$, where $dt \in \{IH, OH, MX\}$ (7)

Here, IH stands for In-House, OH for Out-House, MX for Mixed.

In this use case the value of DT_i domain is mixed (7 and 8).

$DT_i = (MX)$ (8)

$BD_i = (bd_1, bd_2, bd_3)$, where $bd_1 \in \{FE, BE\}$, $bd_2 \in \{OL, RE\}$, $bd_3 \in \{IN, EX\}$ (9)

Here, FE stands for the Front-End and BE for the Back-End parts of the system; OL for OLTP and RE for Reports; IN for Internal users and EX for External users made service.

In this use case the values of BD_i domain are kept as follows: front end development (FE), OLTP support programs (OL), and internal users of the core Java-based system (IN) (9 and 10).

$BD_i = (FE, OL, IN)$ (10)

The complexity group attribute has had significant variation during the observed time period of one year. We need to stress that longer time periods for comparison would increase the risk of other fixed attributes varying. However, a shorter observation time period might not give proper results because some suppliers might have a better learning curve, but could later show a lack of service quality. We propose continuous measurement and immediate internal reporting of poor performance. However, we also stress that actions be taken wisely – only after having a *proper* amount of data during a *proper* amount of time (i.e. once per week/month/quarter, regarding the type of the company, the kind of ecosystems, the amount of concrete work, the acceptable pace of development, etc.).

We classify the complexity C_i of each a_i activity in three levels:

$C_i = (l)$, where $l \in \{1,2,3\}$ (11)

The semantics of our classification is as follows:

1) Level 1 ($C_i = 1$) – low range: from minor changes to the existing code without changes to the data model or component architecture, to medium changes to the existing code that may include data model modifications, but not architecture changes.
2) Level 2 ($C_i = 2$) – medium range: from major changes to the existing code that may include data model modifications and architecture changes, to new application development without significant changes to the surrounding system interconnections.
3) Level 3 ($C_i = 3$) – high range: new system development that may include a number of interconnected applications and significant changes in surrounding system interconnections and replacements.

Since $TD_i, DP_i, DT_i,$ and BD_i values are fixed for this use case, the services activities outcomes were categorized into three service class ranks depending on the complexity of these activities (12).

$$S_i = ((3T, WO, J), C_i, H, MX, (FE, OL, IN)),\ i = 1, 2, 3 \quad (12)$$

The service class attributes value assignments are given for each service class instance, and grouped into predefined services' class ranks (12). All the services in the same service group were then compared.

The selected comparison criteria are based on the hierarchical decomposition of preferences, which operates until the elementary criteria have been reached. In order to compare the quality of the services (the programming services) within a specific rank, we have used the following first, second, and third levels of the hierarchical decomposition of the preferences (starting from global preferences that were recognized at the first level of hierarchical decomposition [Table 2]):

Table 2

Software service preferences hierarchical decomposition

P1 QoS (Quality of Solution/Service)
P11 Maintainability
P121 Changeability
P122 Stability
P123 Testability
P12 Documentation
P13 Performance
P131 Processing time
P132 Throughput
P133 Resource consumption
P14 Reliability
P141 Maturity
P142 Fault tolerance
P143 Recoverability
P15 Usability
P151 Understandability
P152 "Learnability"
P153 Operability
P16 Capability
P17 Installability
P18 Availability
P2 CoS (Cost of Solution/Service)
P21 Fixed Costs
P211 Programming
P212 Licenses

P213 Infrastructure
P22 Variable Costs
P221 Unplanned Development
P222 Unplanned Resources Availability
P223 Travel/Accommodation
P224 Interest Rates
P3 Risks
P31 Strategic
P311 Regulation
P312 Market Position
P313 Shareholders
P32 Operational
P321 Time to market
P322 Unplanned costs
P323 Stakeholders
P323 Retention
P4 Likeability

The group of experts, consisting of IT personnel, PMO, and user representatives (with our support), had been given the task of assigning all weight coefficients and each level correlation logical functions, and providing estimates for each preference during this study in each monthly cycle.

During the first, kick-off, workshop meeting, we defined together (as a team) all the weight coefficients and related logical functions for each hierarchical group level based on our understanding of the importance of each preference estimate and correlation with other preferences within the same group. Each of us gave his/her opinion and we discussed all individual views to come up with the common framework to be used for the measurements and comparisons. We also set a time table for meetings on a monthly basis to discuss the collected production data, to make another 3SI run, and to create an action plan for improvements.

The weight coefficients for the first hierarchical level in this use case are:

$$\omega_1 = 0.4,\ \omega_2 = \omega_3 = 0.25,\ \omega_4 = 0.1 \tag{13}$$

We found that the main features (criteria) of this level of estimation are strongly dependent on each other because higher QoS will, in most cases, produce lower CoS [11], and lower the risks. That is the reason the team gave 40% weight to QoS. We also found that soft features are normally higher with better QoS. This is the reason the r function used in calculation at this level (r_0) is a type of weak conjunction (C-+) (*medium week conjunction* as defined in [34]).

We introduced *Likeability* criteria without purposefully drilling in further (this could be left for further research) to stress the importance that soft criteria be

considered at the highest level of hierarchical decomposition. Based on the data provided, we believe that the core of each successful project or change activity was good communication management. Better communication management gave better results on "Likability" criteria.

We used the following values for weight coefficients at the second hierarchical level (14):

$$\omega_{11} = \omega_{12} = 0.20,\ \omega_{13} = \omega_{14} = 0.15,\ \omega_{15} = \omega_{16} = 0.10,\ \omega_{17} = \omega_{18} = 0.05;$$
$$\omega_{21} = \omega_{22} = 0.50;\ \omega_{31} = 0.70,\ \omega_{32} = 0.30 \quad (14)$$

We used the following values for weight coefficients at the third hierarchical level (15):

$$\omega_{111} = \omega_{112} = 0.35,\ \omega_{113} = 0.30;\ \omega_{131} = \omega_{133} = 0.35,\ \omega_{132} = 0.30;$$
$$\omega_{141} = \omega_{142} = 0.30,\ \omega_{143} = 0.40;\ \omega_{151} = \omega_{152} = 0.20,\ \omega_{153} = 0.60;$$
$$\omega_{211} = 45,\ \omega_{212} = 0.20,\ \omega_{213} = 0.35;$$
$$\omega_{221} = 40,\ \omega_{222} = 0.20,\ \omega_{223} = 0.20,\ \omega_{224} = 0.20;$$
$$\omega_{311} = \omega_{312} = 0.30,\ \omega_{313} = 0.40;$$
$$\omega_{321} = \omega_{323} = 0.30,\ \omega_{322} = \omega_{324} = 0.20 \quad (15)$$

All the above values are discussed among the team members, and are the results of the mutual agreement on importance of each preference within the specific hierarchical group level. The level of precision is influenced by the level of experiences and knowledge of the team. That is the reason why we selected multifunctional team with different experiences in IT projects management, software developments (both from internal and external software services providers) and users of the software systems.

The r function was calculated according to (16). Please note that the value of r is given together with the description of the type of the logical function (i.e. for r_0 we used C-+ function because we wanted to achieve significant level of good estimates for each group criteria at the first hiearchical level), and number of the grouping elements for which it was calculated (i=4 means that we have 4 elements in the group).

$r_0 = -0.235$ (C -+, i=4); $r_1 = r_2 = 1$ (A);

$r_3 = -0.148$ (C -+, i=2);

$r_{11} = r_{13} = r_{14} = r_{15} = 0.573$ (C--, i=3); $r_{21} = r_{22} = 1$ (A);

$r_{31} = -0.208$ (C -+, i=3); $r_{32} = -0.235$ (C -+, i=4). (16)

Please note that the conjunction function was used to stress the need to have good results on all estimations at the same group level.

Since code suppliers normally worked on projects and maintenance activities, we collected data concerning both types of activities. Projects were done in accordance with the predefined project methodology (in this case it was PMI) [37, 38]; all other functional, performance, and bug fixing related changes were considered as maintenance activities and conducted through regular change management procedures. In order to reduce paper work and become more agile, the company only ran as a project those code changes requiring over 15 man-days of work or over 3K Euro in development investment.

Programming activity could be related to non-value adding activities, such as bug fixes (b_i) performed on existing code and new code development (n_i). Regardless of the type of activity programming cost (C) always exists (17).

$$C = \sum_{i=1}^{k} f_n(n_i) + \sum_{i=1}^{l} f_b(b_i),$$

$$\text{where } k \geq 1,\ l \geq 1, \tag{17}$$

The consequences of increased costs due to lower code quality were considered by one or by more sides (directly or indirectly), depending on the contract type between the parties involved. In a time and material type of contract, the sponsor would usually pay for both (r_i and n_i). On the contrary, with fixed term contracts the sponsor, by definition, pays only what was calculated as the amount of work for n_i. A maintenance contract normally covers production issues and new functionalities development. In this case, the sponsor would not be fully aware of the real quality of the deployed code before running deep in the production environment.

Result: By applying 3SI methodology we found that the preserved quality of results and the quality of built mutual relationships among insource and outsource teams on software development activities were correlated (i.e. where the Likability estimate was higher other estimates tended to be higher as well). The figures below (Figure 2 and Figure 3) show the results of team estimates and calculations for the first reporting period (the first month's estimates/the first iteration).

The results show that in this use case overall estimates based on the complexity of the programming activities have not shown any significant difference. All companies gave results that are lower, by up to 5%, as the complexity level grows.

The results show that the lowest estimated difference from the best to the worst supplier in the same complexity class was around 35% (Figure 2 and Figure 3, the difference between provider B and provider D). This significant difference stresses the need for further analysis to discover the root cause of these estimates.

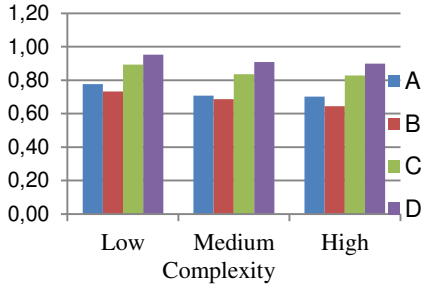


Figure 2
Maintenance activities estimates comparison results per class, per supplier

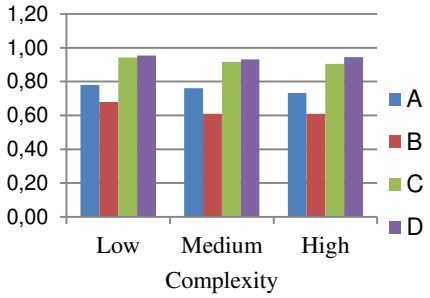


Figure 3
Projects activities estimates comparison results per class, per supplier

In the next step, we identified the major reasons for the differences in the services provided by different suppliers. We started from the most important difference contributor and continued with other major ones (we did not include all reasons in order to avoid gold plating: i.e. to find those 20% of reasons that contribute to 80% of the difference) [26]. The main driver responsible for lowering future costs based on past service performance in dynamic business environments is to have quick (on time) reactions. "What makes measurements so potent is its capacity to instigate informed action – to provide the opportunity for people to engage in the right behaviour at the right time" [39]. After a predefined cycle period of one month, another 3SI cycle was initiated (monthly cycles). For the purpose of this experiment, we repeated these cycles for the period of six months (six iterations). Figure 4 presents the result for six iterations in graphic form.

The cycle periods of 3SI need to be based on the specifics of the business context and content of the software service. We proposed planning time after these periodical assessments to talk to each service provider about the estimates in order to trigger further improvements. In some cases it may work well to have frequent measurements analysis (i.e. on a weekly or monthly basis), but in some cases it

would be advisable to wait for a whole quarter to pass to have a meaningful assessment of performance. In our experiment, after each cycle's data collection and comparison, we scheduled separate interviews with each of the suppliers, during which we presented the findings and asked for feedback. The main result we noticed was *significant improvement in the services of all providers* over time (Figure 4); however, we also noticed/proved that one provider (provider D, Figure 4) had to be replaced (this decision was also based on the measurements, which helped in conducting fair/objective service closure).

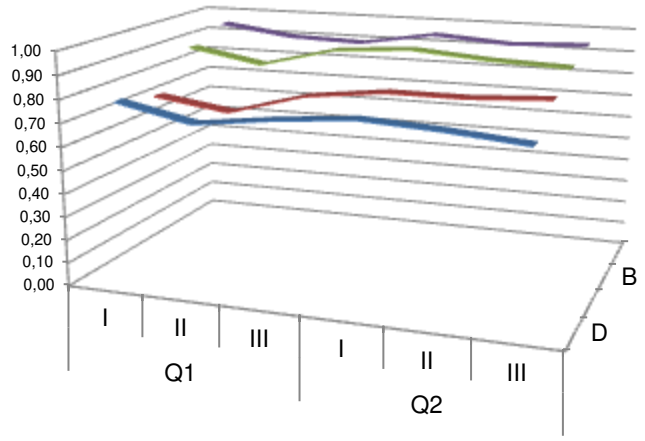


Figure 4
3SI method implementation results for the use case over the given period of time

The cycle of assessments in 3SI could also vary due to the maturity of the business partnership. We suggest starting partnership performance assessments where the learning curve offsets calculation at the beginning (i.e. tolerate up to 10% lower marks due to a learning curve's offsets in the first reporting data sample), and then to continue without offsets into regular comparative analysis and continuous cycles' improvements.

Conclusion

The pressure to give better, cheaper, and faster results in software development is getting stronger as business becomes more and more competitive. A small difference in performance can mean a significant bottom line difference for the company. The fight for knowledge resources does not recognize boundaries. Emerging digital technologies will bring even more stress to the local/domestic office building.

Nowadays, companies rely on some mixture of insourcing and outsourcing for software development activities. In order to manage the risks of using only one outsourcer many companies use two or more outsourcers for the same systems domain. This is mandatory if that domain supports the core company business.

The software services given by code suppliers are devoted to software development projects and/or to maintenance activities on existing software. All these activities represent costs for the company and for the supplier. The goal is to achieve a win-win situation and evenly share the risks regardless of the activity type. This goal, even with detailed contracts in place, is very hard to achieve in practice.

In this paper, we have shown that by using the 3SI methodology it is possible to create estimates for preferences satisfaction for each supplier and to compare them regularly to find a reason why in some areas one supplier may be better than another and vice-versa. The utilization of that knowledge in a regular, systematic way could support the continuous improvement of all services at the same rank.

This case study experiment has been conducted within the live production environment over period of six months, involving a number of people with limited abilities (i.e. with average knowledge and available time that can be taken from regular operational activities). We did our best to optimize utilization of their time, capacity, and availability to simulate a real operational situation. We concentrated our activities on running 3SI in repeating cycles and providing value in improving overall quality of the software services in this specific environment as a main priority.

The fact that comparisons are conducted in periodical cycles could be used to enforce closer communication between parties that could lead to increased quality of the overall services. In some cases, it might lead to calls for contractual expectations that were not seen before or sometimes to even end the partnership. But, this is not the primary goal; this would be an extreme consequence of the new knowledge acquired. The primary goal is to influence and change (if necessary) the parts of the development process that caused an increased number of defects and lower quality of the code.

Since making software is a creative, rather than purely technical, activity with strong human and team interaction, an example from the case study has shown the importance of soft skills and demonstrated certain correlations between Likability criteria and other more quantitative metrics. However, we have also learned that *Likability* estimates can also be improved with the right actions taken.

Further research and implementation: We recommend trying other methodologies (e.g. Squale model) within the same environment and compare the results and the feedbacks from the end-users. In this case, it is absolutely necessary to assess the ratio between costs and achieved benefits in order to reach the primary goal of this paper which is a sufficient gain for obtaining a reasonable price and all this under real constraints primarily related to skills and available time of involved experts.

The utilization of this method in industry can be supported by the creation of a parameterized software solution, which we see as a next practical step. We also envision the need for further research on data analytics based on generated

knowledge base that will support automations in the continual quality improvements (i.e. design of different 3SI *templates* for different industrials' needs).

Further research in soft skills and communications improvements specifically during procurement/selection and suppliers' quality management processes on the complex software projects realizations is also desirable.

References

[1] Boehm B.: Value-based Software Engineering - Reinventing, ACM SIGSOFT Software Engineering Notes, 2003, Vol. 28, No. 2

[2] Gonzalez R., Gasco J., and Liopis J.: Information System Outsourcing: A Literature Analysis, Information & Management, 2006, Vol. 43, No. 7, 821-834

[3] Peppard J. and Ward J.: The Strategic Management of Information System: Building a Digital Strategy, 4th Ed., Wiley, 2016

[4] Chun M. and Mooney J.: "CIO Roles and Responsibilities: Twenty-Five Years of Evolution and Change," Information & Management, 2009, Vol. 46, No. 6, 323-334

[5] Addo-Tenkorang R. and Helo P.: Enterprise Resource Planning (ERP): A Review Literature Report, Proceedings of the World Congress on Engineering and Computer Science 2011 Vol II WCECS 2011, San Francisco, USA October 19-21, 2011

[6] Chin-Sheng C. and Wen-Yau L. M.: A Cloud Computing Platform for ERP Applications, Information & Management, 2015, Vol. 27, No. 7, 127-136

[7] Markovic V., and Maksimovic R.: A Contribution to Continuous Software Services Improvement Based on Six Step Service Improvement Method, IJESEKE, 2012, Vol. 22, No. 4, 1-21

[8] Kan S. E.: Metrics and Models in Software Quality Engineering, 2nd ed., Addison-Wesley, 2003

[9] Betz C.: Architecture & Patterns for IT, Elsevier Inc., 2011

[10] Jones C.: Critical Problems in Software Measurement, Burlington, Mass: Software Productivity Research, 1992

[11] Kuvaja P., Similä J., Krzanik L., Bicego A., Saukkonen S. and Koch G.: Software Process Assessment & Improvement - The Bootstrap Approach, Oxford: Blackwell Publishers, 1994

[12] Pulford K., Kuntzmann-Combelles A. and Shirlaw S.: A Quantitative Approach to Software Management, The AMI Handbook, Wokingham, England: Addison-Wesley, 1996

[13] Karjalainen J., Mäkäräinen M., Komi-Sirviö S., and Seppänen V.: Practical process improvement for embedded real-time software, Quality Engineering, 1996, Vol. 8, No. 4, 565-573

[14] Boehm B. W., Abts C., Brown A. W., Chulani S., Clark B. K., Horowitz E., Madachy R., Reifer D., and Steece B.: Software Costs Estimation with COCOMO II, NJ: Prentice Hall, 2000

[15] Basili V., Daskalntonakis M., and Yacobellis R.: Technology Transfer at Motorola, IEEE Software, 1994, 70-76

[16] Basili V. and F. McGarry.: The Experience Factory: How to Build and Run One, Tutorial TF01, 20th International Conference on Software Engineering (ICSE'98), Kyoto, 1998

[17] Deming E.: Out of the Crisis: Quality, Productivity and Competitive Position, Cambridge University Press, Cambridge, 1982 & 1986

[18] Shewhart W. A., and Deming E. W.: Statistical Method from the Viewpoint of Quality Control, Dover Publications, 1986

[19] Hackman J. R., and Wageman R.: Total Quality Management: Empirical, Conceptual, and Practical Issues, Administrative Science Quarterly, 1995, Vol. 40, No. 2, 309-342

[20] Dean J. W., and Bowen D. E.: Management Theory and Total Quality: Improving Research and Practice through Theory Development, Academy of Management Review, 1994, Vol. 19, No 3, 392-418

[21] Choppin J.: Total Quality Management, What isn't It?, Managing Service Quality, 1995, Vol. 5, No., 147-49

[22] Bergel A., Denier S., Ducasse S., Laval J., Bellinguard F., Vaillergues P., Balmas F., and Mordal-Manet K.,: Squale - software quality enhancement, In Proceedings of the 13th European Conference on Software Maintenance and Reengineering (CSMR 2009), European Projects Track, 2009

[23] Juran J.:Total Quality Management, McGraw-Hill, 2001

[24] Boyd R.: Understanding ITIL Key Process Relationships, Computer Economics, 2007, 1-3

[25] Chrissis M. B., Konrad M., and Shrum S.: CMMI, Guidelines for Process Integration and Process Improvement, Addison-Wesley, 2004

[26] Dion R.: Process Improvement and the Corporate Balance Sheet, IEEE Software, 1993, 28-35

[29] Patton J, Economy P.,: User Story Mapping, O'Reilly Media, 2014

[27] Ciriello R., Richter A., Schwabe G.,: When prototyping meets storytelling: practices and malpractices in innovating software firms, ICSE-SEIP '17 Proceedings of the 39th International Conference on Software Engineering:

Software Engineering in Practice Track, Buenos Aires, Argentina - May 20-28, 2017

[28] Blomkvist J. K., Persson J., and Aberg J.,: Communication through Boundary Objects in Distributed Agile Teams, Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - Chicago 2015, pp. 1875-1884, 2015

[30] Wysocki R. K.: Effective Project Management, 4th ed., Wiley, 2006

[31] Imai M.: Gemba Kaizen: A Commonsense, Low-Cost Approach to Management, McGraw-Hill, 1997

[32] Zsidisin G. A., Jun M., and Adams L. L.: Relationship between Information Technology and Service Quality in the Dual - The Direction Supply Chain: A Case Study Approach, International Journal of Service Industry Management, 2000, Vol. 11, No. 4, 312-328

[33] Salle M.: IT Service Management and IT Governance: Review, Comparative Analysis and Their Impact on Utility Computing, Hewlett-Packard Company, 2004

[34] Dujmovic J. J., and Nagashima H.: LSP method and its use for evaluation of Java IDEs, International Journal of Approximate Reasoning, 2006, Vol. 41, No 1, 3-22

[35] Dujmovic J. J., and Bai H.: Evaluation and Comparison of Search Engines using the LSP Method, ComSIS, 2006, Vol. 3, No. 2, 31-56

[36] Dujmović, J. J.: Continuous Preference Logic for System Evaluation, proceedings of Eurofuse 2005, edited by B. De Baets, J. Fodor, and D. Radojević, pp. 56-80, ISBN 86-7172-022-5, Institute "MihajloPupin", Belgrade, 2005

[37] Culver-Lozo K.: Software Process Iteration on Large Projects: Challenges, Strategies and Experiences, Software Process - Improvement and Practice, 1995, No. 1, 35-45

[38] PMI: Project Management Body of Knowledge (PMBOK Guide), 5th ed., PMI, 2013

[39] Spitzer D. R.:Transforming Performance Measurement, AMACOM, 2007

The X and Y Generations' Characteristics Comparison

Anita Kolnhofer-Derecskei, Regina Zs. Reicher, Agnes Szeghegyi

Óbuda University Keleti Faculty of Business and Management, Institute of Enterprise Management Tavaszmező u. 17, 1084 Budapest, Hungary
derecskei.anita@kgk.uni-obuda.hu, reicher.regina@kgk.uni-obuda.hu, szeghegyi.agnes@kgk.uni-obuda.hu

Abstract: The aim of this paper is to identify generational differences, also similarities among university students in order to realise the reason of generational gap. Data for this study were collected through survey analysis (n=783). Findings indicated significant differences among X and Y generations usage of new technological innovation (i.e. also applications and devices). While Generation Y members (Millennials) are technology and social media addicted, the members of elder generation are in a different life period and due to this they are calmer, and more family oriented. Although generational gap is one of the fanciest management topic it is necessary to highlight that some characteristics are not as socially back-grounded rather than typically age relevant. On the other hand, generation gap is not so sharp the two, there is only a slight shift from X to Y or Z generations. That means the so-called cuspers (between generations' members) should also be taken into consideration.

Keywords: Generation differences; Millennials; Cuspers; X generation

1 Age Groups or Generations?

The foundation for generation research was laid down by American historians William Strauss and Neil Howe in their paper Generations, published in 1991 [2]. Some preferred names for this generation [2] are:

- The new generation of the labour market,
- Generation Y,
- Millennials,
- The next generation,
- Generation Tech,
- Generation.com …

The research has led them to the conclusion that there is a certain pattern in the behaviour of consecutive generations and such a cycle lasts for about 80 years. This lifetime-long generation transition has become much shorter in the case of the recent generations; the quicker the technological innovations are implemented, the more difficult it is to determine the transition between the generations. McCrindley and Wolfinger [6] gave an excellent summary of the characteristics of each generation, but their results were based on the research of Australian generations. Although globalization has been expanding, these limits can be different in different cultures or regions. For example, the authors put the birth of generation Z for 1996; in Hungary, the first members of generation Z have just started their higher education studies. If we use the limit set previously in 1996, we will see later that they are represented in our sample, but the young titans of generation Z are not treated separately from generation Y. (1) Due to lack of an exactly defined turning point, those born in 1996 can also be the elder representative of generation Y; (2) their number of elements as represented in the sample is so low that it does not make any sense to treat them separately.

The sociological research, however, cannot disregard the impact of information technology development. Thus, some research projects distinguish the generations in modern history based on impacts of the digital age. Tapscott [15] basically distinguished 4 generations. The first that he describes in detail is the baby boom generation; they were born between 1946-1964. This generation was named after the explosive growth in number of births after World War II. There are approximately 81 million of them all over the world. The information revolution for them is basically the radio and TV.

Generation X – those born between 1965-1975 according to Tapscott [15] – has arrived at a specific social environment. The number of births has drastically declined, by almost 15%. The unemployment rate is very high; all the jobs have been taken by the members of the older generation. This generation is a very aggressive communicator; they are very strongly media-oriented. They are followed by the Y or Net-generation. Their number is almost as high as the baby boom generation. Their members are extremely efficient in study and work. They almost immediately digest the collected information, share it with their friends and acquaintances, thus helping its interpretation and absorption. The generation Z is the next; they have not entered the labour market yet, they are currently enriching their knowledge in education. They are perfectly fine with all the digital devices and they are almost unable to exist without these devices [14].

First of all, we need to give a correct definition for generation as a niche of a cohort population. The KSH's (Hungarian Central Statistical Office) definition, which has been used since 1970, is the following "Generation is a specific type of population cohort: it means a group of people who were born in the same year. As the members of a generation should live through demographically important events (for example getting a degree, marriage, birth, employment, death, etc.) interlocked this way time and frequency of occurrence of these events are

comparable with factors affecting in time." [4] According to Bordes et al. [1] a generation is a peer group defined by both its demographics and its key life events like passage from birth to adulthood. Shaped by their common history and influences by common icons, events and conditions (forces in the environment) that become reference points for them, which ends in shared values and behaviours.

According to this, the age group born in the same era and thus socialized in a similar environment can be regarded as a generation. Of course, it does not mean that all members of the same age group are identical. The differences of personality and other psychological characteristics make each group colourful.

The environment, in which the person is socialised, is also determinant from the aspect of their system of values. The other important aspect to be highlighted, is that the age differences (developmental psychology) should also be considered, when the generations born in different eras are compared. For example, it has been mentioned in several places that generation Y is ambitious as opposed to the generation X, whose members have become apathetic by now. We strongly believe, however, that the members of generation X were also full of ambition at the same age, when they started to explore the job market. [9]

Reeves and Oh's paper [11] provided a perfect state of art concerning the literature of Generation Differences. They suggested that generation means people born within the same time period (nowadays this period is decreasing); share a common set of characteristics based on historical experiences, economic and social conditions, technological advances and other societal changes. At the same time, they noted that the most popular findings are based on limited data, so these should be treated with a lack of reliability and validity data. But, mainly their paper verifies the above-mentioned KSH's definition, i.e. various generations could be determined by chronological schemes (assign people born in any given year). Opinions of researchers may be different because different geographical areas follow different technical/ecological/business innovation. Cultural differences must be taken in consideration as well, i.e. some chronological schemes and impacts reach the cultures differently and in different times, and we do not take historical differences in consideration or there will be individuals who do not fit the mould.

The paper written by Péter Róbert and Tibor Valuch [12] was chosen as a benchmark. The authors try to describe and classify the political and sociological features of Hungarian history in the last 100 years in terms of generation research, as it is presented in Table 1.

Table 1
Generation division in terms of individual time, historical time and socialization in Hungary [12]

<table>
<tr><td>Historical time</td><td>-1949</td><td>1949-1962</td><td>1963-1989</td><td>1990-1995</td><td>1996-</td></tr>
<tr><td></td><td>Confrontatio, world war, Horthy-era</td><td>The long 1950s</td><td colspan="2">Kádár's consolidation then crisis</td><td>Transformation crisis</td><td>Post-socialism</td></tr>
<tr><td>Individual time</td><td colspan="5">Ageing of generations</td></tr>
<tr><td rowspan="2">Young 1983-1997
Young middle-aged 1963-1981</td><td colspan="3" rowspan="2"></td><td colspan="2"></td><td>Young age socialization</td></tr>
<tr><td>Middle-aged 1948-1962</td><td colspan="2">Family, young-age socialization</td><td>Adult-age socialization</td></tr>
<tr><td>Old -1947</td><td colspan="2">Family, young age – primary socialization</td><td colspan="3">Later, adult age – secondary socialization</td></tr>
</table>

According to this, the members of population regarded as generation Y were born after 1982 and have lived through three historical eras, which could well be distinguished in terms of politics. In our opinion, however, the IT revolution, open borders, globalization and the changes in the education system and attitude have had a much greater impact on this generation than the political and historical structure. Due to the digital revolution, new forms and methods of education have been implemented; new technologies were born, which have presented serious challenges to the traditional educational system. The members of generation Y have lived through these changes. [16] They are 'Baby Boom Echo', this name describes the children of the Boomers. [7] In Hungary, they are the grandchildren of the people of the so-called Ratkó's era.

Managing this period shifts between nations we accept Lancaster and Stillman' [5] opinion. Between two generations there is a slight shift in which group shares preferences, values, attitudes and behaviours of both generations. They are the so-called cuspers. We identified them as those born between 1980-1982. The authors made the distinction between people born on the edges of various generations. They labelled this period between 1975-1980 in the USA. As Reeves and Oh [11] suggested, the existence of cuspers further limits the generalization of generational traits to individuals based on their categorization regarding generation [11, p. 297]. Birth year is only one of the factors to distinguish generations and it is a relatively sensitive factor. We argue that generations are rather shaped by historical heritage and cultural background than chronological data, but this factor could be handled more easily. That is why we use this categorization during our study. Finally, the Hungarian society shows the following picture:

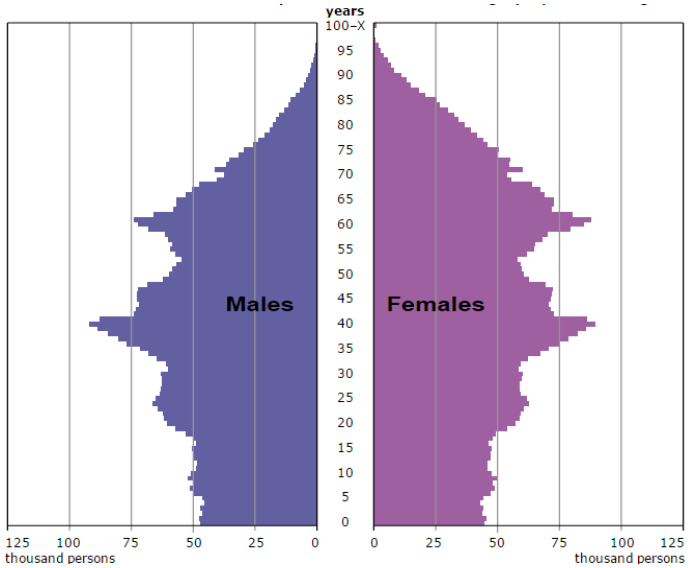


Figure 1
Population number of Hungary by sex and age (1. January 2016 9 830 485 person)

X generation till 37 years old-Cuspers between 36-34 years old-Millennials younger than 33 years old. But, what are the main differences between X generation and Millennials regarding their opinions?

2 Empirical Research

2.1 Research Questions and Methods

We absolutely agree with Reeves and Oh [11] who suggested a distinction between people born on the edges of various generational spans and those caught between two generations by labelling them cuspers. It is difficult to find that salient effect (e.g. political, social or cultural event) which divided the two periods of observed generations. For these, we were interested in the so-called slightly shift. Measuring differences, we had to choose and define factors which could be diverse between generations. Factors and terms used by us were observed earlier, more information can be found in Kolnhofer-Derecskei and Reicher [3].

Our first hypothesis is the following:

H1. There are significant differences between generations' attitudes. (We were interested only in differences so there are no labelled factors.)

We assume that there are main differences between examined generations' groups. Regarding the measured factors there were some, which significantly differ from each other among different generations. These characteristics can be identified with the impact of time and innovation (e.g. using of social media or IT innovations) but, not all of the asked and measured variables will be different. Other characteristics will be similar in the case of all generations, like personal, lifetime period or educational characteristics. Testing our hypothesis, we used Kruskal-Wallis non-parametric test (with sig level. p=0.95) because this method is not so sensitive in distributions and standard deviations of independent samples (compares modes and not means) but it is one of the absolutely reliable statistics. We had to consider the cuspers, due to this, we divided our sample into three layers. Gender differences also consist and influence our way of thinking so *we compared responds regarding gender* (however, this case there was no hypothesis determined we used Mann-Whitney non-parametric hypothesis test).

Secondly, we assumed that factors could be a group of factors. Using the method of clustering we could answer our next research question; namely what kind of characteristics attract different identifiable groups (now generations).

Regarding judgments of subjects, different characteristics could be grouped into three clusters.

With k-hierarchical cluster analysis (in this case exclude cases list wise procedure was used) we could cluster variables (terms) into three clusters (we used k=3 because earlier we have determined three different subsamples, namely generation X, cuspers and Millennials). Using this method, sharp differences will jump out from the data. This procedure attempts to identify relatively homogeneous clusters of variables based on selected characteristics.

Later, based on generated clusters, i.e. variables of grouped characteristics' subjects can be identified because this process also classifies cases (i.e. respondents) and not only variables into the three clusters. Thus, we have generation description based on meaningful clusters and the age category about any of the subjects.

Lastly, but not least, we were able to compare subjects' opinions (clusters) to age categories. Finally, our hypothesis 2 was the following:

H2. Age' and characteristics' clusters are significantly interrelated.

In the last part of our research we used symmetric measures (Cramer's V with 0.95 signif. level) and samples comparison non-parametric method (Friedman k related samples test).

SAMPLE

This study does not examine demographic heterogeneity within millennial cohort because the sample consists of students at Obuda University. This is a typical problem in case of generations' studies, as Reeves and Oh [11] highlighted in case

of Millennials, that the samples were mainly students, focused on people who entered colleges or universities and focused on white collar or intellectual worker careers. However, generation differences mostly cause educational problems before workforce's challenges. But, sooner or later, all the Millennials will enter the labour market. Regarding characterization, we tried to define the research questions from the aspect of the labour market (this is to be discussed and justified in detail in subsequent analyses). When the differences between generations were to be explored, we focused on the comparison made on the basis of labour market features. The methodological background is detailed in Kolnhofer-Derecskei & Reicher [3]. The used listed terms were included in the first question of our question series compiled with the help of Google survey form. Link to the questionnaire:
https://docs.google.com/forms/d/179uhOPyaFpDpvkDAEw5CgPzAzPCbooM04Qhb_cnHYM8/viewform?usp=send_form

The respondents were asked to mark the five most important traits, which they think are typical for the members of their generation. Answering this question was compulsory and the list of characteristics was complemented with an output (text box). The respondents could mark more or less than five traits (the online survey form does not enable to program the given conditions). The completed (and formally as well as semantically validated) questionnaire was sent to all the students of Óbuda University through the Neptun study administration system and the participants in the research could also distribute it. As Rajnai, et al. [10] suggested new generations bring new learning attitudes. Therefore, e-learning is nowadays more than a learning system, it is complete solution and way of communication between students and lecturers. "Many universities and colleges implement e-learning system in their own training programs to satisfy the learners' needs." [10 p. 118] Also, Óbuda University uses Neptun and Moodle systems in order to connect its participants together. That is why we used the Neptun system sharing our survey in the case of measuring students' opinions.

Table 2
Distribution of respondents (head)

Sample	Gender		
	Male	Female	Total
X Generation	**57**	**28**	**85**
Cuspers	**24**	**12**	**36**
Y Generation or Millennials	**403**	**259**	**662**
Total	484	299	783

Since the present paper focuses only on the members of the Y generation, the responses of only those respondents are discussed (after, of course, classification and filtering) whose date of birth was in 1982 or later. The characterization of the sample in terms of demography is summarized in Table 2. The filling ratio in case of students from Obuda University – marked as population – was rather high

Therefore, the sample covers the students of Óbuda University and can be regarded as representative from several aspects.

Regarding our population (i.e. students of Óbuda University) this study is based on selective Millennials and one level of education, rather than national datasets. That caused the over-representation of Millennials. It is important to remember that the data are survey based on self-reports and not behavioural measurements or observations. As we mentioned earlier they will enter the Hungarian corporate world thus, managers and leaders should pay attention to these generation differences and make changes in their leadership styles, practices and tools.

2.2 Hypothesis Examinations Comparing the Opinion of Two Generations

Comparing these results with the responding members of the other generation (about 110 people), we can see the divergence is arising not only from the generation differences, but from age characteristics as well. The following differences are shaped by the comparison of relative frequencies of responses received from students belonging to generations Y, cuspers and X:

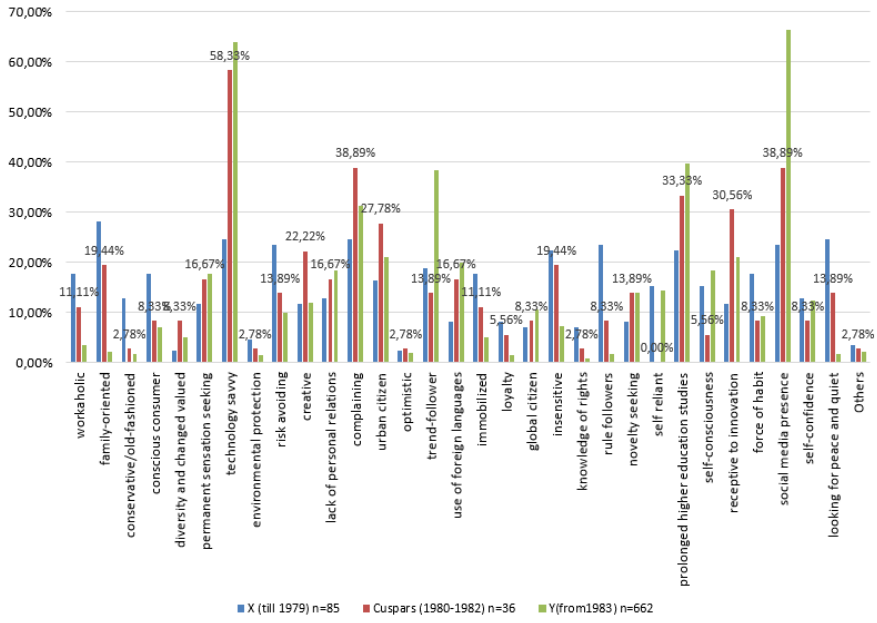


Figure 2
Descriptive statistics of different meanings

For the evaluation procedures, we have used (online) content analysis software and SPSS. Mainly, we used descriptive statistics because most of the responses

were measured on nominal or ordinal scales. Using non-parametric hypothesis test for two independent samples (Kruskal-Wallis test, significance level is 0.05) there were sometimes (n=13) significant differences between the groups. We compared the distribution so in cases of any characteristics there were group diversities. Independent samples median hypothesis test estimated the same results.

Table 3
Summary of our findings

Descriptions	Sig.	Decision
workaholic	,000	Rejected the null hypothesis.
family-oriented	,000	Rejected the null hypothesis.
conservative/old-fashioned	,000	Rejected the null hypothesis.
conscious consumer	,004	Rejected the null hypothesis.
diversity and changed valued	,346	Retain the null hypothesis.
permanent sensation seeking	,377	Retain the null hypothesis.
technology savvy	,000	Rejected the null hypothesis.
environmental protection	,120	Retain the null hypothesis.
risk avoiding	,001	Rejected the null hypothesis.
creative	,197	Retain the null hypothesis.
lack of personal relations	,454	Retain the null hypothesis.
complaining	,267	Retain the null hypothesis.
urban citizen	,361	Retain the null hypothesis.
optimistic	,924	Retain the null hypothesis.
trend-follower	,000	Rejected the null hypothesis.
use of foreign languages	,031	Rejected the null hypothesis.
immobilized	,000	Rejected the null hypothesis.
loyalty	,000	Rejected the null hypothesis.
global citizen	,562	Retain the null hypothesis.
insensitive	,000	Rejected the null hypothesis.
knowledge of rights	,000	Rejected the null hypothesis.
rule followers	,000	Rejected the null hypothesis.
novelty seeking	,334	Retain the null hypothesis.
self reliant	,047	Rejected the null hypothesis.
prolonged higher education studies	,007	Rejected the null hypothesis.
self-consciousness	,120	Retain the null hypothesis.
receptive to innovation	,042	Rejected the null hypothesis.
force of habit	,049	Rejected the null hypothesis.
social media presence	,000	Rejected the null hypothesis.
self-confidence	,756	Retain the null hypothesis.
looking for peace and quiet	,000	Rejected the null hypothesis.
Others	,769	Retain the null hypothesis.

Kruskal-Wallis nonparametric tests algorithm. Asymptotic significances are displayed. The significance level is 0.05 Rejected H0 means: at least one sample is different.

Our first hypothesis (H1) was accepted: we have found significant diverse factors among generations. At the same time, we must highlight that not all the measured characteristics are relevant to generations. Some of them (like creativity or innovation) are independent from the generations' classifications because they are personal traits like optimisms (i.e. born with) or typical general world relevant like urbanism (i.e. globalization problem) or age-relevant (i.e. change with age, education or maturity).

In the second comparison, we divided our sample regarding their gender.

Differences between genders are the following:

Table 4

Group Statistics Mann-Whitney U

gender (grouping variable)		N (Chosen)	Relative (%)	Asymp. Sig (2-tailed)
workaholic	male	23	4,75	0,248
	female	20	6,69	
family-oriented	male	28	5,79	0,954
	female	17	5,69	
conservative/old-fashioned	male	16	3,31	0,619
	female	8	2,68	
conscious consumer	male	39	8,06	0,753
	female	26	8,70	
diversity and changed valued	male	24	4,96	0,971
	female	15	5,02	
permanent sensation seeking	male	86	17,77	0,536
	female	48	16,05	
technology savvy	male	280	57,85	0,228
	female	186	62,21	
environmental protection	male	6	1,24	0,079
	female	9	3,01	
risk avoiding	male	57	11,78	0,863
	female	34	11,37	
creative	male	74	15,29	0,003
	female	24	8,03	
lack of personal relations	male	82	16,94	0,451
	female	57	19,06	

complaining	male	145	29,96	0,465
	female	97	32,44	
urban citizen	male	113	23,35	0,270
	female	50	16,72	
optimistic	male	12	2,48	0,273
	female	4	1,34	
trend-follower	male	165	34,09	0,442
	female	110	36,79	
use of foreign languages	male	88	18,18	0,758
	female	57	19,06	
immobilized	male	33	6,82	0,800
	female	19	6,35	
loyalty	male	12	2,48	0,903
	female	7	2,34	
global citizen	male	48	9,92	0,839
	female	31	10,37	
insensitive	male	55	11,36	0,200
	female	19	6,35	
knowledge of rights	male	8	1,65	0,984
	female	5	1,67	
rule followers	male	24	4,96	0,400
	female	11	3,68	
novelty seeking	male	64	13,22	0,845
	female	41	13,71	
self reliant	male	63	13,02	0,423
	female	45	15,05	
prolonged higher education studies	male	189	39,05	0,270
	female	105	35,12	
self-consciousness	male	84	17,36	0,895
	female	53	17,73	
receptive to innovation	male	86	17,77	0,019
	female	74	24,75	
force of habit	male	49	10,12	0,967
	female	30	10,03	
social media presence	male	275	56,82	0,007
	female	199	66,56	
self-confidence	male	54	11,16	0,231
	female	42	14,05	
looking for peace and quiet	male	31	6,40	0,010
	female	7	2,34	

As can be realised there were four times the significant (sig. level 0.95 $p<0.05$) differences between genders. Creativity and innovation were differently handled, which is quite interesting because these two phenomena depend on each other. Creativity was defined as a personal trait (being creative), and innovation mainly defined as open-minded to innovation. As a conclusion, we can underline that there were not really important differences between the judgments of the two genders.

Our next research question deals with selecting and grouping variables (here characteristics).

First of all, we tried to understand the relationship between variables and for this we used the MDS technique. Multidimensional scaling technique attempts to find the structure in a set of proximity measures between objects, the distances between points in the space matching the given dissimilarities as closely as possible. The result is a least-squares representation of the objects in that low-dimensional space, which, in many cases, will help to further understand data. MDS uses distances between variables, we took Proxscal mode with squared Euclidean distances. The stress and fit measures give an indication of how well the distances in the solution approximate the original distances. Lower stress measures (to a minimum of 0) and higher fit measures (to a maximum of 1) indicate better solutions.

Table 5
Stress and Fit Measures

Dimensionality:	2
Normalized Raw Stress	,04483
Stress-I	,21172[c]
Stress-II	,36124[c]
S-Stress	,08736[d]
Dispersion Accounted For (D.A.F.)	,95517
Tucker's Coefficient of Congruence	,97733

PROXSCAL minimizes Normalized Raw Stress.

c. Optimal scaling factor = 1,047.

d. Optimal scaling factor = ,951.

This method helps us to understand the impact of time instead of grouping characteristics. That is why drawing the cluster centres or borders would not clarify the meaning of our findings. Axes are more important and relevant. Multidimensional scaling helps find the structure in a set of distance measures between our characteristics. Understanding the distances between these points in the space can be interpreted and used to further understand our data.

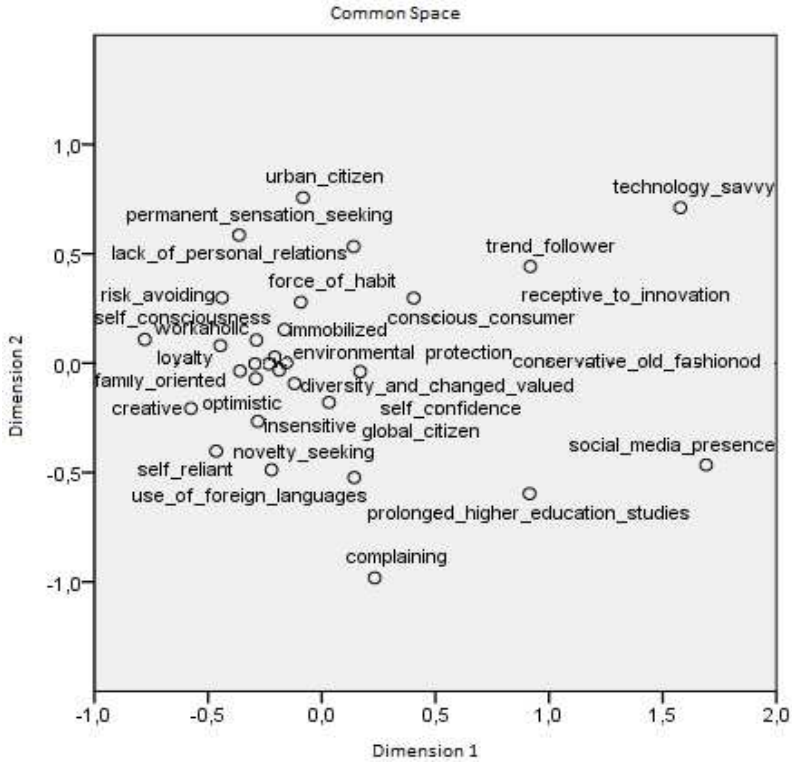


Figure 3
Dimensions

Dimension 1 can be interpreted as a time horizon or generation horizon that means this axis shows the generational gap. Dimension 2 cannot be clearly realised, due to this none of the characteristics on the left side can be interpreted as salient generations' differences. MDS can be used to identify dimensions that describe respondents' perceptions. We found that the time and non-defined horizontal axis define a two-dimensional space, which accounts for the similarities that are reported by our respondents.

As long as MDS draws slight shifts among the characteristics, cluster analysis draws the borders between different groups of variables. t Using cluster analysis makes possible to identify deeper structures within the data. This type of segmentation is able to divide different groups. More specifically, it tries to identify homogeneous clusters' variables.

Clusters of variables are detailed in Table 6. We used three clusters because the original idea based on three subsamples like members of generation X, cuspers and Millennials, but according to the state of art (important literature) we divided

generations regarding their dates of births. Now clustering variables helped us understand connections (here distances) between measured variables.

Table 6
Clusters

Final Cluster Centers

	Cluster		
	1	2	3
workaholic	**,10**	,06	,02
family-oriented	**,14**	,03	,01
conservative/old-fashioned	**,08**	0,00	,01
conscious consumer	**,15**	,10	,02
diversity and changed valued	,05	,06	,04
permanent sensation seeking	,18	,19	,16
technology savvy	,25	,65	**,83**
environmental protection	,03	,02	,01
risk avoiding	**,20**	,04	,09
creative	,15	**,21**	,07
lack of personal relations	,19	,12	,19
complaining	**,39**	,15	,32
urban citizen	**,25**	,13	,21
optimistic	,05	,02	0,00
trend-follower	,10	,31	**,56**
use of foreign languages	,13	**,24**	,21
immobilized	**,12**	,02	,05
loyalty	**,06**	,03	0,00
global citizen	,09	,14	,10
insensitive	**,18**	,02	,06
knowledge of rights	**,05**	0,00	0,00
rule followers	**,11**	,01	,01
novelty seeking	,14	,14	,13
self reliant	,16	,13	,13
prolonged higher education studies	**,44**	,24	,38
self-consciousness	**,21**	,09	,18
receptive to innovation	,01	**1,00**	0,00
force of habit	**,15**	,04	,09
social media presence	,24	,64	**,87**
self-confidence	,13	,11	,12
looking for peace and quiet	**,14**	,01	0,00

*Only the significantly relevant factors (ie. ANOVA p<0.005) were typed in bold.

Cluster 1 can be identified as conservative, old fashioned styled, complaining but self-reliant and conscious group, the members of which are engaged in prolonged studies (if that means Life-Long Learning) this cluster should be generation X. Cluster 2 can be described as conscious consumers who are involved in environment protection and innovation and mainly live in bigger cities. This cluster may be equal to cuspers. The last Cluster, Cluster 3 is creative, speaks many languages, is open for novelty and social media, lastly it is loyal and mainly its members are absolutely technology-addicted. These characteristics are relevant for Millennials. By selecting subjects (cases) based on clusters, we can group them. Cluster memberships were comparing with the original generation order. Comparing these two grouping methods (i.e. characteristics – cluster membership) and age relevant) we found the following connections.

Table 7

generation * Cluster Number of Case Crosstabulation

Count

		Cluster Number of Case			Total
		1	2	3	
generation	X generation	63	9	13	85
	Cusper	20	11	5	36
	Millennial	187	137	338	662
Total		270	157	356	783

Test Statistics[a]

N	783
Chi-Square	270,860
df	1
Asymp. Sig.	,000

a. Friedman Test

Regarding this result, we accepted H2, because we found that our cases' (here subjects) characteristics and clusters ages are linking together significantly (p=0.000 with 0.95 signif. level). The relationship between two variables (i.e. cluster and age grouping) is moderately strong (Cramer's V over 0,3) and significant (p=0.00 with 0.95 signif. level).

Summary

Our findings reflected the state of art concerning various generations' attitudes and habits. H1 was accepted, that means there were different factors between generations.

We absolutely agree with the work of Reeves and Oh [11], some parts of our results based on Hungarian samples verified their findings. Generation X is mainly skeptical and conservative, sometimes regarded as old-fashioned. As long

as Millennials want to be protected as children, Xs are independent and do not like to depend on others for help. For the Millennials the world is open, they speak many foreign languages, are open-minded and accept easily, ambiguous perspectives and opinions, but Xs are aware of diversity and think globally. Millennials are more pragmatic, embrace environmental protection, and they are self-assured and achievement focused. Generation X is more family oriented, so both generations keep in its mind work-life balance. Although members of both generations like to work with the latest technology, the younger one is more technology addicted and technology savvy. They keep in touch with each other through social network media and not really care about personal relationships. That causes a deeply impatient attitude, i.e. prompt feedback whenever they want it.

Summing up, it should be realised that there are no determined differences between these two generations. It was really challenging for us to find the real limits among generations. Regarding our statistical results, no clear border can be drawn that is why it was necessary to use the definition of cuspers. Our results also proved that there is no clear and sharp distinction between people born on the edges of various generational stages. According to our H2, various generations can be distinguished not on the basis of age differences but rather on differently perceived and followed values.

Conclusions

The outcomes of the statistical comparisons were summarized in Table 8, thus revealing the significant differences in the responses received from the members of the two generations.

Table 8

Summary of our findings

Significantly different:	Not significantly different:
- Workaholic * - Family-oriented * - Conservative - Conscious consumer - Technology-addicted (savvy) - Risk-avoiding - Trend follower - Use of foreign languages - Immobilized - Loyal - Insensitive *** - Knowledge of rights - Rule follower - Prolonged higher education studies - Receptive to innovation	- Embracing diversity and changed valued *** - Permanent sensation seeking *** - Environmental protection *** - Creative ** - Lack of personal relations *** - Complaining ** - Urban citizen *** - Optimistic ** - Global citizen ** - Novelty seeking ** - Assertiveness ** - Self-consciousness ** - Self-confidence **

- Force of habit - Social media presence - Looking for peace and quite ***	

In our opinion, there are some traits among the above, which are age specific (marked with *) and some which can be led back to personality differences (marked with **), and finally some others, which are problems of our era independently from the age or lifetime period (marked with ***). It would be worth, however, to explore these differences with further review of literature.

Some popular articles suggest that Millennials are a “want it” and “want it now” generation, but we cannot judge a book by its cover [8]. When we (i.e. X gen members) were fresh graduates and young, did we not want the same? Having good pay and benefits, rapid successes, work/life balance and of course challenging work? Although the career expectations of the Millennials are manifested differently in their career ladders, their needs and motivations are not really altered from their earlier imaginations. Mostly the companies try to suit their structures, organizations and styles to their employees. Nicholas [7] gave a good description of the workplaces of various generations. The new generation will not respect authority; respect is not based on title or position; all their colleagues must earn their respect. Finally, “It has been suggested that many of the career goals and expectations among Millennials are ‘‘supersized,’’ unrealistic, and disconnected between reward and performance.” [8 pp. 282]

Or are we responsible for these “problems”. According to Sacks [13] they are said to have “helicopter parents” who hover over them (the over-involved Boomer parents). Other typical outcomes of this hovering behaviour are – among others - the Pan Peter syndrome or Mama-hotel and Papa-bank.

We agree with Takács, et al. [14 p. 30], who suggested that "Robotics is becoming a mainstream phenomenon, entering all areas of our lives. In addition to cutting-edge research and development, robotics is becoming equally important in the classroom and home education." Robotics education has grown tremendously in significance, one reason for that can be the needs and wants of this new generation (i.e. Millennials) whose members are saliently technological savvy. [14]

As we have referred to it in several points of the present paper, this question has been discussed as part of a larger, comprehensive research. We had dual objectives; not only the self-introduction of generation Y has been presented with the help of projection technique, but – in terms of education methodology - we also conducted research with direct and active involvement of students, where the students were both the target group and the active researcher.

Our outcomes have further strengthened the research concerning characterization of the two generations but it was a novelty that the members of the generations themselves evaluated their own generation. Fine shift could be observed among any generation, the members of these shifts are called cuspers. During this

research, we have explored cuspers, and the results of our sample layers underline the continuous changes.

It is highly important to understand that there are a lot of features in the characteristics of generations, which are more age-specific and to a lesser extent due to socialization. The authors of the present article have also shifted their focus to family as they were getting older. The discrepancies between generations are often due and strengthened by the different communication of the two generations. As long as a generation is being educated and served by the previous generation, this education will be responsible for any issues with the new generation.

Acknowledgement

We are thankful for the suggestions of 2014/2015 spring semester Research Methodology Students. This paper supported by the ÚNKP-16-2/I. New national excellence program of the ministry of human capacities. We are thankful for the suggestions of the two anonym reviewers, that helped us to make more clarified this paper.

References

[1] Borges, N. J. – Manuel, R. S. – Elam, C. L. – Jones, B. J. (2006): Comparing Millennial and Generation X Medical Students at One Medical School. Academic Medicine. Vol. 81/6 571-576

[2] Howe, N. – Strauss, W. (2009): Millennials Rising: The Next Great Generation. Knopf Doubleday Publishing Group

[3] Kolnhofer-Derecskei, A. – Reicher, R. (2016): Genyus – how do millennials judge themselves? In: Maksym Bezpartochnyi (ed) Theoretical, methodological and practical foundations of human resources management. Riga: ISMA University, pp. 25-32

[4] KSH (2015): Módszertani dokumentáció/fogalmak - Generáció, http://www.ksh.hu/apps/meta.objektum?p_lang=HU&p_ot_id=200&p_obj_id=646 (2015. 03.)

[5] Lancaster, L. C. – Stillman, D. (2003): When Generations Collide. Who they are. Why they clash. How to solve the generation puzzle at work. New York Collins Business. (Core references)

[6] McCrindle, M. –Wolfinger, E. (2009): The ABC of XYZ. Understanding the Global Generations UNSW Press

[7] Nicholas, A. G. (2009): Generational Perceptions: Workers And Consumers. Journal of Business & Economics Research. 7/10, 47-52

[8] Ng, E., S., W. – Schweitzer, L. – Lyons, S. T. (2010): New Generation, Great Expectations: A Field Study of Millennial Generation. Journal of Business Psychology. 25. 281-292

[9] Pais, E. R. (2013): Alapvetések a Z generáció tudománykommunikációjához, PTE

[10] Rajnai Z., Nguyen Huu Phuoc Dai ; Duong Van Thinh (2016): Learning attitude in XXI century Published in: Applied Machine Intelligence and Informatics (SAMI), 2016 IEEE 14th International Symposium on p.: 115 - 119

[11] Reeves, T. C. – Oh, E. (2008): Generational differences. In Jonassen, D. - Spector, M. J. Driscoll, M., - Merrill, M. D. - Jeroen van Merrienboer (ed) Handbook of Research on Educational Communications and Technology: A Project of the Association for Educational Communications and Technology. pp. 295-303

[12] Róbert, P. – Valuch, T. (2013): Generációk a történelemben és a társadalomban. Generációs politikai attitűdök és részvételi minták történeti-szociológiai megközelítésben. Politikatudományi Szemle XXII/4. pp. 116-139

[13] Sacks, D. (2006). Scenes from the culture clash. Fast Company, 102, 72-77

[14] Takács Á., Eigner Gy., Kovács L., Rudas I., Haidegger T. (2016): Teacher's Kit: Development, Usability and Communities of Modular Robotic Kits for Classroom Education IEEE Robotics & Automation Magazine 23:(2) pp. 30-39

[15] Tapscott D. (2009): Grown Up Digital: How the Net Generation is Changing Your World, McGrawHill, New York

[16] Todd, R. J. (2008): The digital Revolution and Web 2.0 – Impacts and challenges on school libraries and the learning environment http://www.nalis.gov.tt/LinkClick.aspx?fileticket=o3Db0RG%2Bijc%3D&tabid=64 (2015. 03.)

Phase-locked Loop Based on Integrating Scanning Conversion

Maksim Dudkin, Aleksandr Nesterov, Aleksandr Shishkov

Department of Electric Drives, South Ural State University (national research university); prospekt Lenina 76, 454080, Chelyabinsk, Russia; dudkinmm@susu.ru, nesterovas@susu.ru, shishkovan@susu.ru

Abstract: A new principle for developing a phase-locked loop (PLL) based on integrating scan conversion is proposed. The PLL is a nonlinear pulse system based on the closed structure of an integrating scanning converter (ISC) operating in the mode of external synchronization with circuit voltage frequency. The block diagrams, the waveform diagram and frequency synchronization conditions for integrating scanning converter switch by outer signals – bipolar rectangle and harmonic, are provided. The recommendations for choosing the frequency ratio of self-oscillations of the scanning converter and circuit voltage frequency were tested. It is shown that the integrating PLL fully adapts to amplitude instability and circuit voltage frequency within the range of ±50% and higher. Unlike the existing PLLs, the proposed system is a first-order adaptive filter, whose bandwidth is automatically tuned to the amplitude and frequency of circuit voltage. This property of the integrating PLL allows suppressing higher harmonics in the supply circuit with frequency instability within the range of ±50% and higher. The system's high noise immunity to commutation distortions of circuit voltage is also shown. These results show the practicability of using it in semiconductor converter control systems, which receive power from an independent low-power supply system, for example, diesel or wind power generating plants. The limits of the dynamic parameters of this system for gradual change of amplitude and circuit voltage frequency have been determined. A three-phase reverse thyristor rectifier is used as an example of a crossed synchronization flow-chart. The recommendations on choosing the parameters of the integrating PLL and also the results of the practical application of the synchronizing unit as a part of semiconductor converter control systems are provided.

Keywords: control system; integrator; scanning converter; semiconductor converter; phase-locked loop

1 Introduction

Modern electric energy systems performed on a stationary basis (and on an autonomous basis as well) are characterized by a high level of distortion with parameters which are difficult to predict, and which often go beyond distortion

standards and which act as destabilizing factors in the performance of not only the semiconductor inverters, but the control system in general [1-3]. This is why development of a semiconductor converter (SC) control system that is capable of full adaptation to changes in circuit parameters (amplitude, frequency, higher harmonics, commutation and impulse distortions) in addition to high noise immunity is an important issue directed towards improving the performance reliability of the entire electro-technical equipment complexes of industrial plants. Creating such a system is important for application not only to development in the SC area, but to working equipment, in need of overhaul, as well.

The most important element of any SC control system is the synchronizing units (SU) since the accuracy of the control system's synchronization with the supply circuit depends on its operation. In the majority of cases, the SU is out of the closed control loop of the SC control system, and therefore any destabilizing effect of circuit voltage leads to a deviation of its output characteristics. If the supply power is high (minimum distortion and stable circuit parameters), it is usually possible to obtain a stable undistorted synchronization signal without any problems. However, in the case of low-power supply, it is quite a challenge to obtain an undistorted synchronization signal since it must exactly match the phase of the first harmonic of circuit voltage.

The first synchronizing units were designed in accordance with the open circuit principle and were represented by cascade connection of the smoothing filter (for example, aperiodic first-order and relay element with switching thresholds which have a relative asymmetry level of zero) [4]. The disadvantage of this type of SUs is apparent – the defined angle of synchronization fundamentally changes as the amplitude is alternated [5], which affects the SC characteristics in general, and can lead to emergency shutdown.

The application of band-pass filters [6], which detect the useful component of the frequency of zero-phase shift circuit voltage, can be useful when the SC operates in a low-power circuit. However, the main disadvantage of these devices is that they generate noncharacteristic harmonics, which cause problems of harmonic instability. Therefore, we had to discard the use of band-pass filters in the SC control systems.

The development of synchronizing circuitry in view of the above-mentioned disadvantages contributed to the design of closed synchronizing units based on the principle of a phase-locked loop [7-9].

The simplest PLL system consists of an error detector based on a multiplier, a low-pass filter (LPF), a proportional-integral controller (PI), and a voltage controlled oscillator (VCO) (Fig. 1a) [7]. In this PLL, the error signal u_{ERR} in phase and frequency is generated as a result of multiplying the synchronizing voltage $X_S(t) = t \cdot \sin(\omega_1 t + \theta_1)$ and the feedback signal $t \cdot \cos(\omega_2 t + \theta_2)$ and is equal to

$$u_{ERR} = 0{,}5 \cdot \sin[(\omega_1 - \omega_2) \cdot t + (\theta_1 - \theta_2)] + 0{,}5 \cdot \sin[(\omega_1 + \omega_2) \cdot t + (\theta_1 + \theta_2)] \quad (1)$$

The steady-state synchronizing voltage $X_S(t)$ is equal to the feedback signal. In this case, $\omega_1 = \omega_2$ and $\theta_1 = \theta_2$, therefore the first term in expression (1) equals zero. The second steady-state term represents an uncontrolled component with the frequency 2ω. To detect the error signal of the constant voltage u_F and to eliminate the uncontrolled variable component, a low-pass filter (LPF) with a transfer function $W(p) = 1/(T_F p + 1)$ is used. The output signal of the PI-controller $\Delta\omega$ $\Delta\omega$ and the initial frequency ω_0 are used to modulate the frequency and phase of the rotation angle θ, as well as the formation of $\sin(\omega t)$ and $\cos(\omega t)$ at the output of the sinusoidal oscillator (SO).

The disadvantage of the conventional PLL is that it does not provide precise synchronization with the synchronizing signal $X_S(t)$ when its frequency varies within a wide frequency range, for example, when the frequency deviates from the steady-state value within the range of ±50% and higher. This is explained by the fact that the LPF cutoff frequency and the PI-controller parameters have fixed values and are tuned to the initial frequency ω_0.

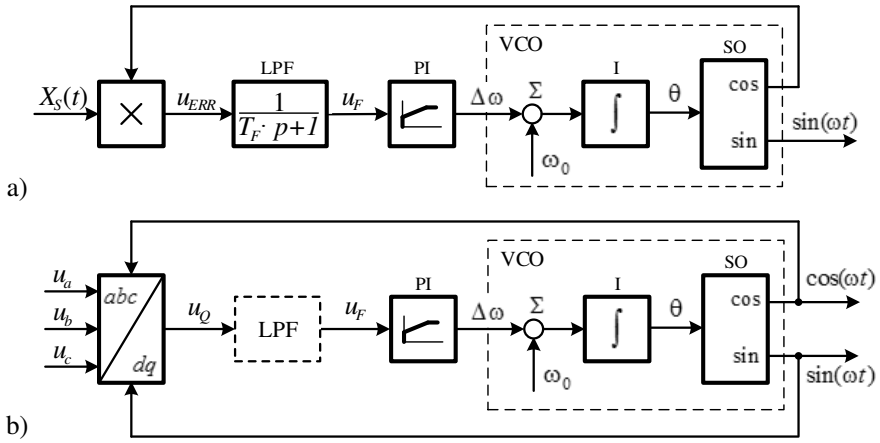


Figure 1

PLL block diagram: conventional (a); based on the *p-q* theory (b)

The PLL system based on the *p-q* theory (Fig. 1b) eliminates this disadvantage [8-9]. The system functions on the basis of the well-known property of the Park-Gorev transform [10], which converts a three-phase coordinate system *abc* into a rectangular synchronous *dq* one. Thus, an error signal $u_{ERR} = u_Q$, corresponding to the difference in frequency and phase between the supply circuit voltage and the feedback signals $\sin(\omega t)$ and $\cos(\omega t)$, is generated at the output of the coordinate conversion unit. In other respects, the composition of blocks and the operation principle of the PLL circuit (Fig. 1b) completely coincide with those of the conventional PLL (Fig. 1a).

The main difference between the conventional PLL and the PLL based on the p-q theory is that there is a higher 2nd harmonic in the error signal u_{ERR} of the conventional PLL and the closed loop must allow for an LPF under normal operating conditions. At the same time, the PLL error signal u_Q (Fig. 1b) lacks such component of alternating voltage, so an LPF is not required.

However, if the power bus voltage contains higher harmonics, for example, the 3rd one, then the error signal u_Q will contain the 2nd voltage harmonic. In order to suppress it, an LPF (Fig. 1b) has to be integrated into the PLL system based on the p-q theory. And if the supply circuit frequency is unstable within the wide frequency range (±50% and higher), it will be necessary to reconfigure the LPF and the PI-controller so that the PLL can achieve high synchronizing accuracy and stability of its closed loop.

The above-mentioned problem is solved in the PLL based on integrating scan conversion [11]. Hereinafter, this system will be simply referred to as "integrating PLL".

The original contribution of the article is that it proposes a new principle for developing a PLL on the basis of integrating scan conversion, which differs from the existing PLL structures by the ability to suppress higher harmonics in the power supply circuit with frequency instability being within the range of ±50% and above.

Numerous studies conducted by the authors showed that applying the methods of integrating scanning conversion for designing both synchronizing units and semiconductor inverter control systems on the whole is one of the most effective ways of improving their noise immunity, the static and dynamic accuracy [12-16].

The next sections describe an integrating phase-locked loop [11], which combines high noise immunity to circuit distortions (higher harmonics and commutation notches) and an ability to adapt to the oscillation of amplitude and power voltage frequency in a wide range.

2 Static and Dynamic Characteristics of the Integrating PLL

In the core of the integrating PLL lies a continuous structure of an integrating scanning converter with pulse-frequency-width modulation (PFWM) [13, 16], which consists of adders Σ1–Σ2, an integrator I and a relay element (RE) with a hysteresis loop (fig. 1a) of 'zero' relative symmetry and which works in outer synchronization with the supply-line voltage mode [13].

A synchronization signal $X_S(t)$ (a supply-line voltage), which has a sine wave form (Fig. 2b) or a rectangle of bipolar impulses with an average value of zero,

needs to be supplied to the integrating scanning converter (ISC) input for conversion of ISC from the PFWM mode to the forced switch mode, when the pulse-width modulation is being implemented. As a result, the alternating impulses $Y(t)$ with the supply-line voltage frequency of $(T_S)^{-1} = (T_0)^{-1}$, which are shifted by the relatively $X_S(t)$ signal to the angle of synchronization α_S (Fig. 2b, c), reach a steady state on the ISC output. The output integrator signal is proximal by form to the harmonic signal $X_S(t)$. Switching of RE is being executed in achieving a switching threshold $\pm b$ EL by an evolvent of integrator $Y_I(t)$ (Fig. 2c).

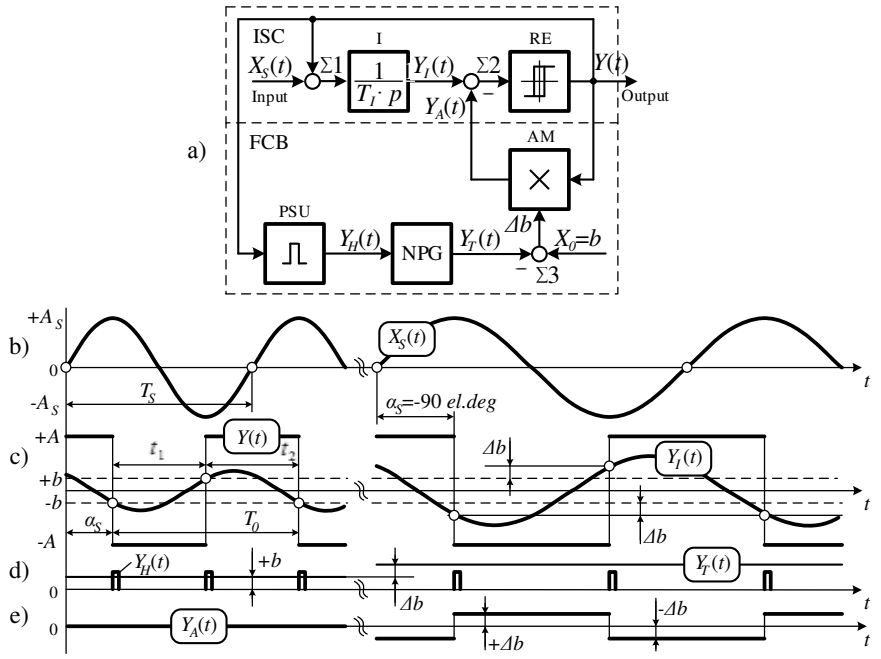


Figure 2

Block diagram of the integrating PLL (a) and waveform diagrams of its signals (b-e)

Conditions for ISC synchronization by rectangle bipolar pulses and harmonic signal are described by the following inequalities:

$$\bar{A}_S \geq |\bar{T}_0 - 1| \tag{2}$$

$$\bar{A}_S \geq 0{,}5\pi \cdot |\bar{T}_0 - 1| \tag{3}$$

In these inequalities $\bar{A}_S = |A_S/A|$ is the scaled amplitude of the A_S synchronization signal (degree of synchronization); $\bar{T}_0 = T_0/T_S$ is the scaled value of the ISC self-auto-oscillation $T_0 = 4\overline{b} \cdot T_I$, when $A_S = 0$, in relation to T_S the period of the synchronization signal $X_S(t)$; $\bar{b} = |b/A|$ is the scaled value of the switching threshold of the RE; $\pm A$ is the amplitude of RE output pulses; T_I is the ISC time constant.

Analysis of equations (2), (3) leads to the following conclusions:

– while $\bar{T}_0 \ll 1{,}0$, the depth of ISC synchronization by rectangle pulses should be chosen considering the $\bar{A}_S \geq 1{,}0$ constraint, and $\bar{A}_S \geq \pi/2$ for the harmonic signal. Otherwise, ISC switches to the self-oscillation mode with the frequency $f = (1 - \bar{A}_S^2)/4\bar{b} \cdot T_I$;

– while $\bar{T}_0 > 1{,}0$, the level of $\bar{A}_S$, required for ISC synchronization by rectangle pulses, as well as by harmonic signal, rises steeply. It is explained by the fact that in order to synchronize the ISC of low undamped frequency $(T_0)^{-1}$ by high frequency synchronization signal with the frequency of $(T_S)^{-1}$, $\bar{A}_S$ needs to be increased in such a way that the equation for second-volt area module $(S_S = |A_S \cdot T_S|) > (S = |A \cdot T_0|)$ is exactly satisfied. Otherwise, the pulse frequency $(T_0)^{-1}$on the ISC output is less than the frequency$(T_S)^{-1}$ of $X_S(t)$ synchronization signal;

– while $\bar{T}_0 = 1{,}0$, the $\bar{A}_S$ is approaching zero, which is why choosing the self-oscillation frequency $(T_0)^{-1}$ of ISC equal to the frequency $(T_S)^{-1}$ of synchronization influence $X_S(t)$ is appropriate.

The phase shift α_S, between the synchronization influence $X_S(t)$ and input impulse $Y(t)$ (Fig. 2b, c), depends on the scaled value of self-oscillation period $\bar{T}_0 = T_0/T_S$ and the depth of synchronization $\bar{A}_S$

$$\alpha_S = -90el.deg. \cdot \left[\!\left[1 + \frac{\bar{T}_0 - 1}{\bar{A}_S}\right]\!\right]. \quad (4)$$

Equation (4) shows that while $\bar{T}_0 = 1{,}0$, the ISC self-oscillation frequency $(\bar{T}_0)^{-1} = 1/4\overline{b} \cdot T_I$ equals the signal frequency $(T_S)^{-1}$ of synchronization $X_S(t)$ and $\bar{A}_S$ has any value, the synchronization angle α_S always stays constant and equals –90 el. deg. This gives the integrating PLL the properties for adaptation to instability of circuit voltage amplitude within the range of ±50 % and higher, which provides the basis of the space of static state $\Delta\bar{\alpha}_S = f(\Delta\bar{A}_S, \bar{A}_S)$ (Fig. 3a) derived from *MatLab+Simulink* simulation of the integrating PLL with the change in synchronization depth within the range $0{,}25 \leq \bar{A}_S \leq 10{,}0$. In this equation $\Delta\bar{\alpha}_S = (\alpha_S{}^* / \alpha_S) - 1$ is the scaled error of synchronization angle $\alpha_S{}^*$ in relation to the anterior synchronization angle α_S as the outer distortion is absent, and which possesses a value of -90 el. deg; $\Delta\bar{A}_S = (A_S / A_{S.N}) - 1$, $\Delta\bar{f}_S = (f_S / f_{S.N}) - 1$ are the scaled errors of amplitude A_S and frequency f_S of supply-line voltage $X_S(t)$ in relation to their scaled values $A_{S.N}$ and $f_{S.N}$ respectively.

As the frequency $\Delta\bar{f}_S$ and the synchronization signal degree $\bar{A}_S$ change, the area $\Delta\bar{\alpha}_S = f(\Delta\bar{f}_S, \bar{A}_S)$ takes on substantially non-linear pattern (Fig. 3b). Thus, with the increase of $\Delta\bar{f}_S$, the error $\Delta\bar{\alpha}_S$ increases due to the violation of the equality between the ISC self-oscillation frequency $(T_0)^{-1}$ and the supply-line voltage frequency $(T_S)^{-1}$ and according to (4), the synchronization angle α_S deviates from the scaled value equal to -90 el.deg. This is one of the disadvantages of the integrating PLL.

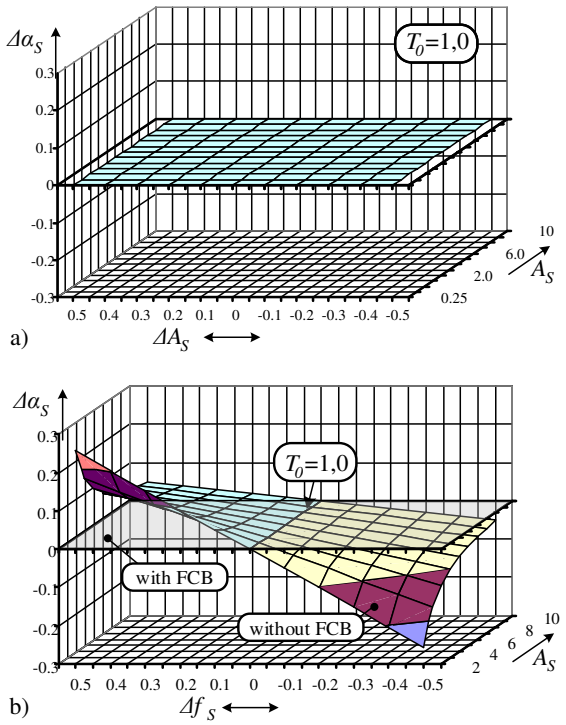


Figure 3

The space of static state $\Delta\bar{\alpha}_S = f(\Delta\bar{A}_S, \bar{A}_S)$ (a) and $\Delta\bar{\alpha}_S = f(\Delta\bar{f}_S, \bar{A}_S)$ (b) of the integrating PLL for outer synchronization by harmonic signal

For elimination of the described disadvantage, a frequency correction block (FCB) is included in the PLL circuit (Fig. 2a). It consists of an amplitude modulator (AM), an adder Σ3, a voltage supply $X_0 = b$, a phase-to-voltage converter (PVC) and a narrow-pulse generator NPG, which starts at the rising and trailing impulse flanks of the ISC output.

At the scaled circuit voltage frequency the signal on the PVC output $Y_T(t)$ equals the ER switching threshold b (Fig. 2d) in absolute value. This is why $Y_A(t)$ equals zero on the AM output (Fig. 2e).

As the circuit voltage frequency $X_S(t)$ oscillates from the scaled value (Fig. 1b), the signal $Y_T(t)$ increases to $(b + \Delta b)$ level on the PVC output (Fig. 2d). An alternating signal $Y_A(t)$ with an amplitude Δb and period that equals period T_0 of the ER output signals, is formed on the AM output (Fig. 2 c, e). This changes the ER switching threshold $|b|$ by $|\Delta b|$, satisfying the equation $T_S = T_0 = 4\overline{b \cdot T_I}$ (Fig. 2 c). As a result, in a steady-state mode, the synchronization angle α_S (between the supply-line voltage $X_S(t)$ and the ISC output pulse) stays -90 el.deg. (Fig. 2 b, c). That is why the synchronization angle error $\Delta\bar{\alpha}_S$ equals zero. This proves the static state area (Fig. 3 b). In these circumstances, when the frequency

correction block is added to the PLL circuit, the circuit voltage frequency range of motion significantly expands from -0.9 to 3.0 (or from 5 to 200 Hz in absolute numbers at the scaled circuit voltage frequency $f_{S.N} = 50Hz$) as the synchronization signal level is $0{,}25 \le \bar{A}_S \le 10{,}0$.

To determine the degree of dependence of the ISC dynamic characteristics on the parameters of the synchronizing signal $X_S(t)$, we applied *MatLab + Simulink* software to make the Bode magnitude plot $k = f(\bar{F})$ of the integrating PLL system synchronized by both rectangular and harmonic signals with different values of synchronization depth (Fig. 4). An alternating harmonic signal $X_A(t) = A_A \cdot sin(2\pi f_A t)$ with a constant amplitude $\bar{A}_A = \left|\frac{A_A}{A}\right| = 0{,}1$ and frequency f_A was applied to the ISC input that was also exposed to the synchronizing effect $X_S(t)$. The central component Y_0 of the ISC output pulses was isolated using a digital filter that implements the algorithm $Y_0 = A \cdot (t_2 - t_1)/T_0$. Here, t_1, t_2, T_0 are the time intervals and period of the ISC output pulses (Fig. 2c), and A is its amplitude. In this case, the input signal level $X_A(t)$ and the average value Y_0 of the ISC output pulses get uniquely connected only after the repetition period T_0 of the pulses $Y(t)$ ends.

Fig. 4 uses the following notations: $k = Y_{0.m}/A_A$ is the ISC transmission coefficient; $\bar{F} = f_A/f_S$ is the scaled frequency of the harmonic signal $X_A(t)$ with respect to the frequency f_S of the synchronizing effect $X_S(t)$.

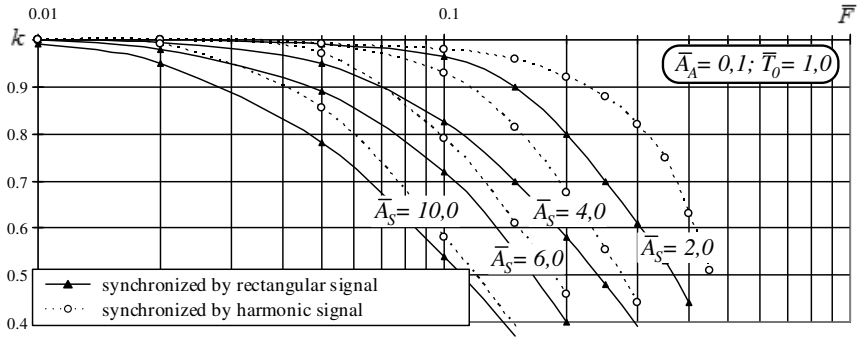


Figure 4

Bode magnitude plot of the integrating PLL with different values of synchronization depth

Analysis of the Bode magnitude plot (Fig. 4) allows us to conclude [17] that the integrating PLL's dynamic properties are close to ones of the first-order aperiodic link $W(p) = 1/(T_E p + 1)$ with the time constants equivalent $T_E \approx 0{,}25\bar{A}_S \cdot T_S$ (5) for synchronization by rectangle pulses and $T_E \approx (\pi \cdot \bar{A}_S \cdot T_S)/16$ (6) for synchronization by harmonic signal.

However, it needs to be considered that this "linearization" is valid for only the $\bar{F} = f_A/f_S \le 0{,}5$ area of input frequency impact, which is derived from the Kotelnikov theorem [18]. If the frequency is higher, the scanning converter, being

a pulse system, switches to the slow sampling of dynamic input signal mode [13, 19].

– the integrating PLL is an adaptive filter whose parameters readjust into the functions of the amplitude $\bar{A}_S$ and the frequency ${T_S}^{-1}$ of the synchronization signal (of circuit voltage). As a result, the integrating PLL synchronizes with the circuit voltage with high accuracy (the synchronization angle error $\Delta\bar{\alpha}_S$ tends to zero) even if there are higher harmonics in the supply voltage and frequency instability of the synchronizing signal within the range of ±50% and higher. This is confirmed by the results of the integrating PLL simulation when the synchronizing signal $X_S(t)$ has the third lowest-frequency voltage harmonic with the amplitude $\bar{A}_{S(3)} = A_{S(3)}/A_{S(1)}$ that is equal to 30% of the main component amplitude $A_{S(1)}$ of the circuit voltage with two frequency values $f_S = 50\ Hz$ (Fig. 5a) and $f_S = 10\ Hz$ (Fig. 5b) of the supply voltage.

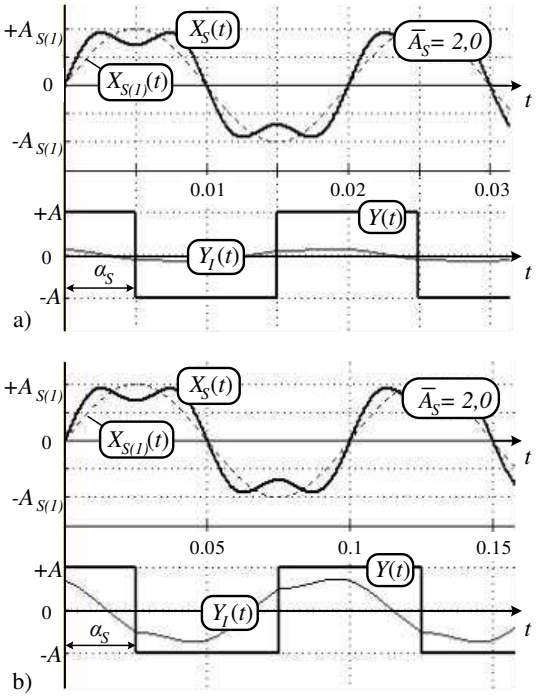


Figure 5

Waveform diagrams of the integrating PLL with the third harmonic in the synchronizing signal and the change in the supply voltage frequency: $f_S = 50\ Hz$ (a) and $f_S = 10\ Hz$ (b)

Fig. 5 shows that the 3rd harmonic is completely suppressed by the integrating PLL, and the synchronization angle α_S is equal to minus 90 el. deg. It is noteworthy that, other things being equal, a synchronization error occurs in the existing PLL systems (Fig. 1) because of the fixed values of the low-pass filter

bandpass and the PI-controller parameters that are tuned to the nominal frequency of the synchronizing voltage;

– representation of the integrating PLL in the form of a first-order aperiodic filter indicates its high noise immunity to commutation and pulse distortions of circuit voltage. This is confirmed by the results of the integrating PLL system simulation when there are commutation notches in circuit voltage that reach the level of $\gamma_C = 25$ el. deg. (Fig. 6a). The notches were created by a three-phase bridge thyristor rectifier of current when it operated for a low-power circuit. In this case, the synchronization angle error was $\Delta\bar{\alpha}_S = 0$ ($\alpha_S = -90$ el. deg.), since commutation notches "1" and "2", "3" and "4" mutually compensated each other in the time intervals t_1 and t_2, or areas S_1', S_1'' and S_2', S_2'' in subintervals t_1', t_1'' and t_2', t_2'' equal each other (Fig. 6b).

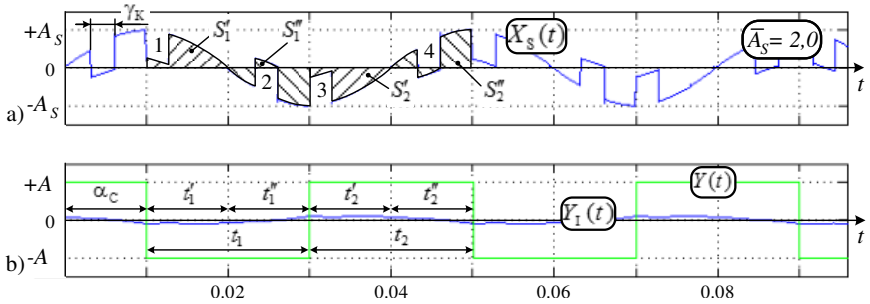


Figure 6

Waveform diagrams of the integrating PLL at commutation notches in circuit voltage ($\gamma_C =$ 25 el. deg.)

In a real industrial environment the over-voltage amplitude and the circuit voltage frequency dips, generally, appear not instantly, but after a certain time interval, which depends on the loading conditions and the type of load included in the circuit. This is the reason why the analysis of the integrating PLL operation in the conditions of the smooth change in amplitude (Fig. 7a) and the frequency (Fig. 7b) of the circuit voltage appears to be of interest.

In Fig. 7 the following notations have been introduced: $\Delta\bar{A}_S$ is the scaled error of an actual circuit voltage amplitude A_S in relation to its rated value $A_{S.N}$; $\bar{S}_U = \Delta A_S / A_{S.N}$ is the scaled velocity of change in amplitude ΔA_S in one period T_S of circuit voltage in relation to its rated amplitude $A_{S.N}$; $S = \frac{(T_S)^{-1} - (T_{S.N})^{-1}}{1sec}$ is the absolute velocity of change in the circuit voltage frequency $(T_S)^{-1}$ in the set time interval t_3 equal to 1 second.

The results of the integrating PLL research have shown that in the dynamic operations the allowed velocity of change in amplitude $\bar{S}_U$ in one circuit voltage period and the frequency S_f makes ±11 % and ±9 Hz/s in 1 second in the conditions when the maximum synchronization angle inclination $\Delta\alpha_S$ does not

transcend ±2 el.deg., and the depth of synchronization is chosen out of the range $\pi/2 \leq \bar{A}_S \leq 8{,}0$ (Fig. 8).

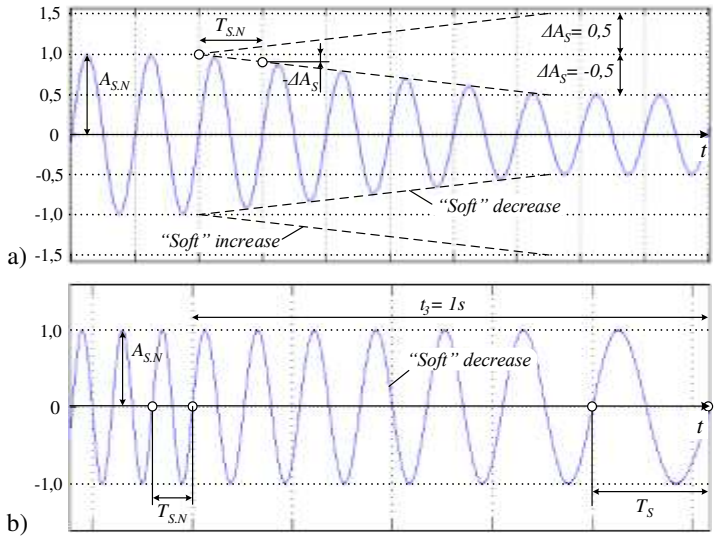


Figure 7

Test impact for analysis of the dynamic characteristics of the integrating PLL at the smooth changes of amplitude (a) and frequency (b) of the circuit voltage

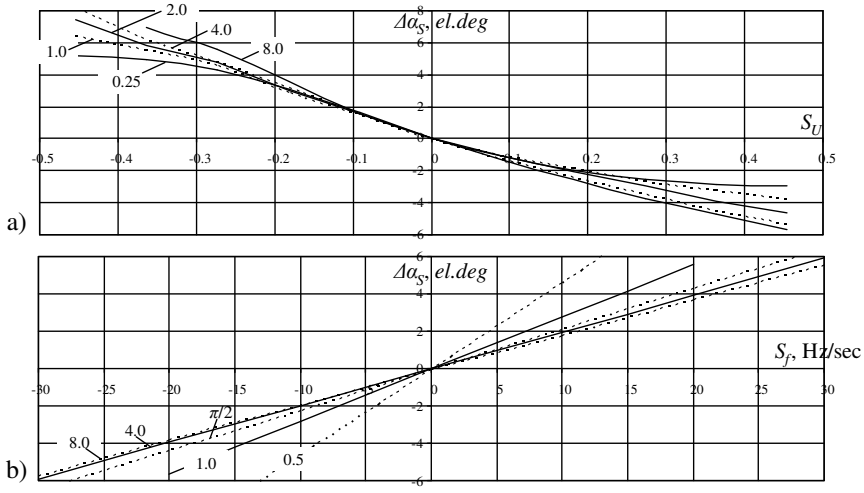


Figure 8

Graphs of dependencies of the absolute synchronization angle error $\Delta\alpha_S = f(\Delta\bar{S}_U)$ (a) and $\Delta\alpha = f(S_f)$ (b) for integrating PLL at different depth of synchronization $\Delta\bar{A}_S$

The optimal level $\bar{A}_S$ is considered to meet the range of $\pi/2 \leq \bar{A}_S \leq 4{,}0$, when the speed of response and the immunity to noise are compromised.

The above results suggest the efficiency of using the integrating PLL in the control systems driven by SC and supplied by autonomous low-power supply systems with a high-level of distortion (higher harmonics and commutation notches) and unstable circuit voltage parameters as well.

Occurrence of the phase shift α_S (equal to -90 el.deg) between the synchronization signal $X_S(t)$ and the output pulses $Y(t)$ (Fig. 2b, c) of the integrating PLL (Fig. 2a) involves an untraditional set-up of the synchronization circuit of the SC control system (CS). One of the set-up options, for example, is the method of CS channel cross synchronization that can be applied in a three-phase thyristor rectifier with a synchronous multichannel control system (Fig. 9a).

Here the PLL-A output sync-pulse (with the duration of 180 el.deg.) outruns phase A by 90 el.deg. (Fig. 9b). Herein the rising flank of this pulse coincides with the natural commutation point of phase C. As a result, the synchronization of the phase C control channel can be done by the PLL-A output sync-pulse. The synchronization of other SC control channels can be done the same way (Fig. 9a).

In order to get the sync-pulses to coincide with the momentum of phase voltages going through the zero point, PLL-A, PLL-B and PLL-C should be synchronized from linear voltages *AB*, *BC* and *CA*, which outrun the phase ones *A*, *B*, *C* by 30 el.deg.

The integrating PLL (Fig. 2a) can be used in the control systems for active voltage (current) rectifiers, active power filters, matrix frequency converters and other SC. These engineering solutions are considered in more detail in the paper [20].

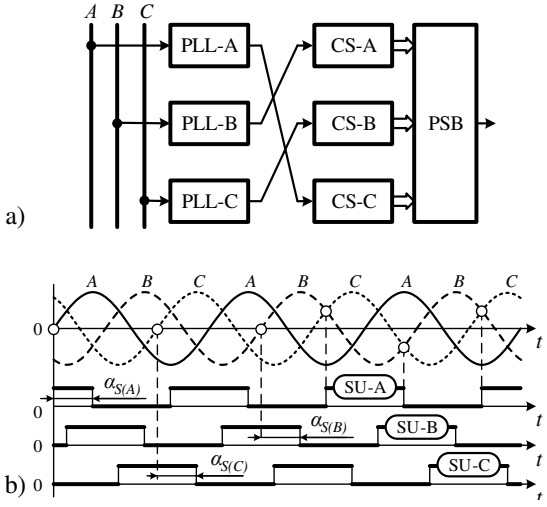


Figure 9

Block diagram (a) and the signal waveform diagram (b) of cross synchronization system of the semiconductor inverter (PLL-A, PLL-B, PLL-C – PLL of phase channels A, B, C; CS-A, CS-B, CS-C – control system for phase channels A, B, C; PSB – power semiconductor block)

Fig. 10 shows the waveform chart of the circuit voltage and the output signal of the integrating PLL system, which functions in the cross-strapping mode (Fig. 9a) in the reverse thyristor rectifier for an electric drive of a direct current control system. The experiment took place on the diesel-generator plant of a drilling rig of the VI group. The total load of the plant was about 90%. Trouble-free performance of an entire technological complex was provided, despite the abnormally high-level of the circuit voltage distortion. This is achieved due to the ISC closed structure and an integrator in its direct control channel. With respect to the synchronizing effect, the system represents a first-order aperiodic filter with an equivalent time constant, which readjusts depending on the circuit voltage parameters.

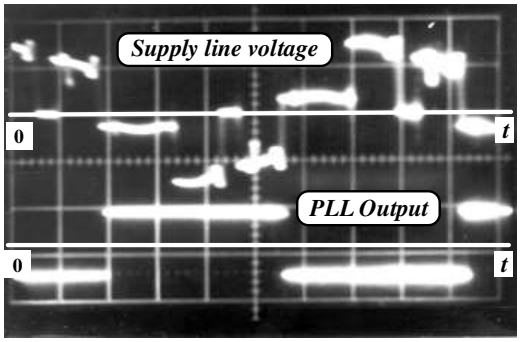


Figure 10
Waveform chart of the circuit voltage and the PLL system output signal in co-operation of several thyristor rectifiers for a low-power supply system

The PLL diagram, showed in Fig. 2a, was also tested on 84 AC thyristor regulators for the soft start of asynchronous electric drives [21] at the JSC "Chelyabinsk Tube-Rolling Plant", and showed high noise immunity and operational reliability.

Conclusions

An analysis of the static and dynamic characteristics of the integrating PLL, based on the integrating scanning conversion has been made. The PLL is a nonlinear impulse system based on the locked structure of an integrating scanning converter (SC) operating in the mode of external synchronization with circuit voltage frequency. It has been determined that for the SC the frequency of self-oscillation should be chosen equal to the scaled synchronization signal frequency. It excludes the drop out of outer synchronization mode of the device in the face of significant circuit voltage failures. At the frequency deviation of the supply voltage the equation listed above is satisfied due to the inclusion of the frequency correction block in the PLL system.

It has been shown that the integrating PLL system fully adapts to the instability of the amplitude and frequency of the circuit voltage within the range ±50% and

higher. Unlike the existing PLL systems, the proposed system is a first-order aperiodic filter $W(p) = 1/(T_E p + 1)$ with the equivalent time constant $T_E \approx (\pi \cdot \bar{A}_S \cdot T_S)/16$ and $T_E \approx 0{,}25\bar{A}_S \cdot T_S$ for the harmonic synchronization signal and the rectangular one) that automatically adjusts depending on the amplitude and frequency of the circuit voltage. This property of the integrating PLL system allows suppressing higher harmonics in the power supply system, while frequency instability is within ±50% and higher. Moreover, the proposed PLL system has high noise immunity to commutation distortions of the circuit voltage. This is achieved due to the closed SC structure and the presence of an integrator in its direct control channel.

Basing on the studies, we have formulated the recommendations on the choice of the integrating PLL system parameters. The optimal level of synchronization depth $\bar{A}_S$ is the range $\pi/2 \leq \bar{A}_S \leq 4{,}0$ where speed and noise immunity are compromised. In dynamic operating modes, the permissible change rate of the amplitude in one circuit voltage period is ±11%, and that of the frequency per one second is ±9 Hz/s, provided that the maximum deviation synchronization angle does not exceed ±2 el. deg, and the synchronization depth is selected according to the condition $\pi/2 \leq \bar{A}_S \leq 8{,}0$.

We have proposed a cross method for synchronizing the channels of SC control systems based on the integrating PLL. We have provided the results of the practical operation of the cross-strapping synchronization being a part of control systems of the reverse thyristor rectifier for a direct current electric drive and thyristor voltage regulator for soft start of asynchronous electric motors, which proved its high noise immunity and operation reliability.

Thus, we recommend that the proposed integrating PLL be used in SC control systems that are supplied by an autonomous low-power supply system with a high-level of distortions (higher harmonics and commutation notches) and unstable circuit voltage parameters.

Acknowledgement

The work was supported by Act 211 Government of the Russian Federation, contract № 02.A03.21.0011.

References

[1] Laszlo Tihanyi: Electromagnetic Compatibility in Power Electronics, Newnes, New York, 1995, 403 p.

[2] GOST 32144-2013. Electric energy. Electromagnetic compatibility of technical means. The quality norm of electric energy in general-purpose power supply systems, 2013

[3] Henry W. Ott: Noise reduction techniques in electronic systems, John Wiley & Sons, New York, 1988, 872 p.

[4] Marian P. Kazmierkowski, R. Krishman and Frede Blaabjerg: Control power electronics, Academic Press, USA, 2002, 518 p.

[5] M. M. Dudkin and L. I. Tsytovich: Adaptive integrators to control power valve converters, Russian Electrical Engineering, Vol. 86, No. 12, 2015, pp. 681-685

[6] J. Jacques, F. P. Dawson and R. Boriert: A tracking bandpass filter for the rejection of power system disturbances, EPE'9 35th European Conf. on Power Electronics and Applications, Vol. 4, No. 377, 1993, pp. 47-52

[7] Dan H. Wolaver: Phase locked loop circuit design, Prentice Hall, Inc., 1991, 272 p.

[8] F. Liccardo, P. Marino and G. Raimondo: Robust and Fast Three-Phase PLL Tracking System, IEEE Transactions On Industrial Electronics, Vol. 58, No. 1, 2011, pp. 221-231

[9] P. Rodriguez, J. Pou, J. Bergas, J. I. Candela, R. P. Burgos and D. Boroyevich: Decoupled Double Synchronous Reference Frame PLL for Power Converters Control, IEEE Transactions On Power Electronics, Vol. 22, No. 2, 2007, pp. 584-592

[10] B. K. Bose: Power Electronics and Variable Frequency Drives, IEEE Press, New York, 1996, 822 p.

[11] M. M. Dudkin, RF Patent 2,513,024, 2012.

[12] M. M. Dudkin, L. I. Tsytovich and A. S. Nesterov: Adaptive units and control systems of power semiconductor converters on the basis of integrating scanning conversion, Procedia Engineering, Vol. 129, 2015, pp. 933-939

[13] M. M. Dudkin, L. I. Tsytovich and O.G. Brylina: Spectral characteristics of scanning converters with pulse-width and pulse-frequency-width modulations, Russian Electrical Engineering, Vol. 84, No. 10, 2013, pp. 549-555

[14] L. I. Tsytovich, O. G. Brylina, M. M. Dudkin, R. M. Rakhmatulin and A. V. Tyugaev: Analog-digital converter with integrating in-phase amplitude-frequency-pulse modulation for switched drive systems, Russian Electrical Engineering, Vol. 84, No. 10, 2013, pp. 244-249

[15] L. I. Tsytovich, M. M. Dudkin, O. G. Brylina and A. V. Tyugaev: Integrating pulse-number ADC with high temporal and temperature stability of characteristics, Automatic Control and Computer Sciences, Vol. 49, No. 2, 2015, pp. 103-109

[16] L. I. Tsytovich, M. M. Dudkin, O. G. Brylina and A. V. Tyugaev: Dynamics of communication lines with pulse-width and pulse-frequency-

width data carriers. Theory and practice, Russian Electrical Engineering, Vol. 85, No. 4, 2014, pp. 183-191

[17] L. I. Tsytovich, O. G. Brylina, M. M. Dudkin and R. M. Rakhmatulin: Adaptive interval-code integrating synchronization of control systems for power converters, Russian Electrical Engineering, Vol. 84, No. 3, 2013, pp. 122-128

[18] C. Bissell: Vladimir Aleksandrovich Kotelnikov: Pioneer of the Sampling Theorem, Cryptography, Optimal Detection, Planetary Mapping ... History of Communications, Communications Magazine, IEEE, No. 10, 2009, pp. 24-32

[19] L. I. Tsytovich and M. M. Dudkin: Dynamics of scanning converters with sampling of instantaneous values of the control signal and various modulation patterns, Russian Electrical Engineering, Vol. 85, No. 10, 2014, pp. 619-627

[20] M. M. Dudkin: Energy-saving technologies in test stands with the use of single-phase reversible converters, Bulletin of SUSU. Series "Power engineering", Vol. 13, No. 1, 2013, pp. 5-18

[21] M. M. Dudkin: Thyristor voltage regulator with adaptive integrating control system for smooth start-up of asynchronous electric motors, Bulletin of SUSU. Series "Power engineering", Vol. 14, No. 2, 2014, pp. 36-43

[22] J. Dudrik and J. Oetter: High-Frequency Soft-Switching DC-DC Converters for Voltage and Current DC Power Sources, Acta Polytechnica Hungarica,Vol. 4, No. 2, 2007, pp. 29-46

[23] J-S. Guan et al.: Breast Tumor Computer-aided Diagnosis Using Self-Validation Cerebellar Model Neural Networks. Acta Polytechnica Hungarica, Vol. 13, No. 4, 2016, pp. 39-52

[24] R.-E. Precup, S. Preitl: PI-Fuzzy Controllers for Integral Plants to Ensure Robust Stability, Journal Information Sciences, Vol. 177, 2007, pp. 4410-4429

[25] F. Tahri, A. Tahri, A. Allali and S. Flazi: The Digital Self-Tuning Control of Step a Down DC-DC Converter, Acta Polytechnica Hungarica,Vol. 9, No. 6, 2012, pp. 49-64

[26] H. K. Lam, F. H. F. Leung and P. K. S. Tam: Stable and Robust Fuzzy Control for Uncertain Nonlinear Systems, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Vol. 30, No. 6, 2000, pp. 825-840

Modelling the System of Receiving Quick Answers for e-Government Services: Study for the Crime Domain in the Republic of Serbia

Vojkan Nikolić

Ministry of Interior Republic of Serbia, Kneza Miloša 101, 11000 Belgrade, Serbia, vojkan.nikolic@mup.gov.rs

Branko Markoski

Technical Faculty "Mihajlo Pupin", University of Novi Sad, Đure Đakovića bb, 23000 Zrenjanin, Serbia, markoni@uns.ac.rs

Kristijan Kuk, Dragan Ranđelović, Petar Čisar

Academy of Criminalistic and Police Studies, Cara Dušana 196, 11080 Zemun, Serbia, kristijan.kuk@kpa.edu.rs, dragan.randjelovic@kpa.edu.rs, petar.cisar@kpa.edu.rs

Abstract: Today, e-Governments services operate quick answer systems in many different domains in order to help citizens receive answers to their questions at any time and within a very short time. The automatic mapping of relevant documents stands out as an important application for automatic questions-documents classification strategies. This paper presents a contribution to the identification concepts in text comprehension in unstructured documents as an important step towards clarifying the role of explicit concepts in information retrieval in general. The most common representational scheme in text categorization is the Bag of Words approach when the dictionary, as incorporating background knowledge, is large. The authors introduce a new approach to create concept-based text representations, and apply it to a text categorization collection to create predefined classes in the case of short text document analysis problem. Further in this paper, a classification-based algorithm for questions matching topic modelling is presented. The paper also describes the weighting of concepts that present terms with high frequency of occurrence in questions is based on their similarity relevance in the predefined classes of documents. The results of the experiment within the criminal law domain in the present case study show that the performance of concept-based text representations has proven to be satisfactory in the case when there is no dictionary for this domain.

Keywords: e-Government services; criminal law domain; natural language processing (NLP); Bag of Concept (BOC) approach

1 Introduction

The reform and modernisation of the public sector, based on a wide application of information-communication technologies (ICT) is considered to be one of the key elements in furthering development of the information society in the Republic of Serbia. The trends of development of many e-Government services in numerous countries in the world indicate the necessity of quick answer systems and e-Government services in the Republic of Serbia, as well. When performing many of the text mining and natural language processing tasks, it is necessary for all forms of a given word with the same meaning to be of identical form. On the whole, many documents that are available to e-Government services tend to be not structured, thus it is very difficult to isolate some forms (knowledge) from them.

Question Answering (QA) systems are an integral part of a multitude of e-Government services. E-Government represents the usage of information retrieval and extracting technologies in order to improve the activities of public sector organizations so that their online services make public information available "anytime, anywhere" to citizens. In cooperation with the government organizations, these QA systems usually aim to compile an annual list of commonly asked questions. In that way, each organization can develop their own online database of related responses especially given that the majority of citizens need assistance from e-Government services for a specific domain providing specific answers to citizen queries. However, the following problem arises: these systems need to hold all the adequate or similar answers able to answer the citizen's questions within the system. It is often the case that the system does not have prepared answer sets. Therefore, the existing Frequently Asked Questions (FAQ) website pages may play a crucial role in the collection of all possible answers.

To extract keywords in Serbian, the authors used a e-Government support component, the so-called ADVANSE [1]. The process that enables the automatic recognition of key words within the text document is known as the "Bag of Words" or "Bag of Concepts", it is a process in which small collections of recognized words and phrases are extracted. In order for this process to function, it is necessary to use certain algorithms so as to reduce the text based on important characteristics. Rapid Automatic Keyword Extraction (RAKE) is one of the simplest algorithms used to extract key words [2]. There are also cases, when the word corpus in the text cannot be recognized making it thus not possible to extract key words. Usually, in text mining set of co-occurring words within a given window is computing by N-grams of texts. The basic point of N-grams is that they capture the language structure from the statistical point of view. In such an instance, the N-gram is used to create a "Bag of N-grams," which represent the specific terms of the given text or document. This method is often used when searching texts written in natural languages. Key words in documents that have similar content or where the same language text corpus is used, can be extracted

by using algorithms for indexing TF-IDF (term frequency - inverse document frequency) [3].

For their search service based on tokenization, counting and normalization, Apache Corporation and Microsoft have created several natural languages processing technology. Following experimental features of the search service API showed that Apache Lucene analysers are faster, but that the Microsoft analysers have advanced capabilities (lemmatization, word decompounding and entity recognition). In the present case, the authors developed the QA system using the open-source search server based on the Lucene Java search library.

The paper is organized as follows. Following this introduction, the second section of the paper elaborates on the related works. The third section presents the basic theoretical principles of text mining techniques with special emphasis on short texts and the lexical normalization (stemming) process based N-gram analysis. Also, in the same section the report in criminal law documents is given along with mapping paragraph members of the Criminal Code of the Republic of Serbia as documents for response answering question approach. The subsequent section describes a framework for the proposed quick answers, which is implemented on the web service related to the e-Government for the criminal law domain. Further details about the components of the architecture of this proposed system are given and the functionalities with the developed algorithm for question Classification are shown, too. Finally, in the experimental results section, the authors report on the successful implementation of the proposed system as well as the recommender for the best similarity measures for finding similarity between short documents in Serbian.

2 Related Work

Taking into account the existing solutions which depend on e-Government requirements, Šimić and others [1] firstly considered the problems identified in the existing solutions in the Republic of Serbia and then offered their own hybrid approach solution: multi-layer based document clustering based on the model of fuzzy concepts and using different measures of text similarity. The aim is to decrease the time of receiving a response from subjects (institutions) of public administration with a minimum (or completely without) any civil clerks' involvement, so the authors offered a new approach to facilitate the optimization and automation of using advanced methods and techniques of retrieving information. This paper presents the ADVanced ANSwering Engine solution (ADVANSE) for a wide range of e-Government services. The most important results of the ADVANSE project are reflected in the quality of e-Government services, the quick response time to the citizens' requests, the innovative uses of the available content and the restructured relationship between civil clerks and

citizens. In particular, the emphasis is on the efficiency and flexibility of the response.

The present group of authors proposes focusing on testing in different conditions and improving the ability of adaptation in the next research phases. One of the objectives pursued in this work is to find solutions for the functioning of such a system in multilingual environments and increasing content complexity concerning grammar and dictionaries of different languages regardless of the area of use. In that sense, different strategies are proposed. A particular challenge is the functioning of e-Government services in different domains. Specific domains use special dictionaries, so it is necessary to use specialized techniques to find similarities. Further, a qualitative improvement of processing the given document is also required. A possible way to obtain the answer is the tagging of certain parts of the document instead of labelling the entire document.

Marovac et al. [4] conducted an analysis of the texts written in different languages and with different linguistic rules. It is especially complicated to analyse the texts written in Serbian given that the Serbian language has two alphabets in use, Cyrillic and Latin, and is also one of the languages with a rich morphology. The use of linguistic resources (text corpus of contemporary Serbian language, morphological dictionaries, stop-words, dictionary of abbreviations, etc.) aiming to obtain a qualitative analysis of a natural language becomes a considerable challenge.

The use of N-gram analysis contributed in significant results achieved without the use of extensive lexical content or analysis of texts in Serbian carried out without using the morphological vocabulary. Special attention is paid to the algorithm for extracting keywords (N-gram) shown in more detail below.

According to [4], the algorithm should be developed like clustering keywords according to their frequencies in the text or parts of the text or clustering keywords (N-gram) that are frequent in comparison to those less frequent ones, i.e. the ones having higher chi-square value. The size of the chi-square statistic is a misleading indicator of the extent of the relationship between two words. In this way, relationship of keywords bases can be extracted.

Stanković et al. [5] also deal with a model that leads to better quality of extensive information retrieval in text document and databases using a bag-of-words for pre-indexing and naming these objects. Each document contained in the database represents the summary report consisting of meta data (domains, keywords, summary and geographic location). The bag-of-words, resulting from these metadata using morphological dictionaries, are objects identified when using the rule-based system. The work of these authors has been valuable in the process of pre-indexing documents (information retrieval). The ranking of the documents downloaded is based on several measures TF-IDF. It can therefore be concluded that the development results obtained in such a way compared to the ones obtained

without the use of pre-indexing showed a great progress regarding an average of accurate measurement.

Kolomiyets et al. [6] represent the Question Answering method as a comprehensive approach that provides a more qualitative way for information retrieval. This approach is actually a system of queries and documents in relation to the possible functions of search in order to find an answer. This research discusses general questions contained in a complex architecture with increasing complexity and the level of frequency of questions and information objects. These authors represent here a method of how natural language roots are reduced on keyword for search while knowledge databases and resources, obtained from natural language questions and answers, are made intelligible.

Wenyin et al. [7] presented a user-interactive question-answering platform named BuyAns, a kind of online community which primarily uses a scheme for answering questions among users in this way creating a special pattern-based user interface (UI) for asking and answering. This pattern is based on the pooling and storage of pairs obtained in the question-answer relationship. Such a system promotes a C2C business model for exchanging and commercializing different types of knowledge. It can be used as a special kind of incentive approach to knowledge acquisition. In operation, this system uses templates and recognizes accurate answers to specific repeated questions.

3 Theoretical Background

3.1 Lexical Normalization

In general, optimization involves identifying the extreme values of objective functions given in a particular domain which combine different types of objective functions and different types of domains. In mathematics, statistics, empirical science and information technology, optimization is the selection of the best element (with regard to some criteria) from a given set of available alternatives [8]. When referring to the problem of optimization, one means the minimum and maximum chosen among real functions which are given, where systematic changes of input values (with regard to functions) are predetermined and defined.

An example of a text mining technique is given in [9], as the most commonly applied technique in the field of e-Government, citizen participation and e-democracy [10]. The QA application (in regard to text mining) enables choosing/finding the best answer for questioning and answering in the context of e-Government system with multiple search text options. These techniques imply text categorization/classification, text clustering, the concept of object extraction, sentiment analysis, document summarization, etc. which help in the determination

of relations within the collection of a large amount of text data using specific forms and methods. Given that text classification/document categorization is the process of assigning predefined categories of free text documents providing an overview of the concepts within the collection of documents, this solution has found a wide range of practical usage [11]. A great number of current approaches represent the text as a list (vector) of words, the so-called "Bag-of-Words" (BoW). In fact, based on the statistical analysis of certain factors (such as frequency and distribution within the collection of words), and without taking the word order and grammar into account, BoW creates a series of words and phrases recognized as important.

The lexical features are most commonly used in the QA and summarization communities. One of the robust lexical feature approaches in statistical models is the use of the N-gram model. The N-gram analysis determines the frequency of different N-grams in a text. N-grams of texts are extensively used in text mining and natural language processing tasks. The N-gram overlap measures the overlapping word sequences between the candidate document sentence and the query sentence. The recall-based N-gram scores for a sentence P using the following formula [12] are as follows:

$$NgramScore(P) = max_i \left(max_j \, Ngram(s_i, q_j) \right) \tag{1}$$

$$Ngram(S, Q) = \frac{\sum gram_n \in S^{Count_{match}(gram_n)}}{\sum gram_n \in S^{Count(gram_n)}}, \tag{2}$$

where n stands for the length of the N-gram ("n = 1, 2, 3, 4"), and $Count_{match}$ ($gram_n$) is the number of N-grams co-occurring in the query and the candidate sentence, q_j is the j-th sentence in the query pool, and s_i is the i-th sentence in the sentence pool of sentence P. In general, an N-gram is a subsequence (substring) of size n from a given sequence (string). An N-grams of the size 1, 2 or 3 are referred to as anuni-gram, bi-gram or tri-gram, respectively.

Lexical normalization (stemming) is the process of reducing words to their roots. This is a necessary step in analysing textual content quickly. Stemming is the process of removing all of the prefixes and suffixes of the given words so as to produce the stems or the root [13]. The main objective of the stemming process is to remove all possible affixes (prefixes and suffixes) and thus reduce the word to its stem. N-gram analysis can be used as an alternative for finding words whose root is the same. This method identifies groups of words where the words in the group are small syntactic variants of one another and collects only the common word stem per group. For example, "development", "developed" and "developing" will be all treated as "develop". In order to obtain more effective results in the selection of the number n, the suffix are collected for important words in Serbian language [14]. Based on this, it was found that the highest percentage (92.70%) were extensions of under 4 letters. Also, in the research study of the paper by the authors Marovac et al. [4], the 4-gram analysis is shown

as it presents the best results for Serbian. For further research in this paper, therefore, 4-gram analysis will be performed as the reference algorithm for normalizing words in the documents in Serbian language.

3.2 Mapping Paragraph Members of the Criminal Code as Short Texts to Concepts

The BoW framework requires faster performance and adaptation of new words [16], especially during the search of short texts, as well as a simple online application. Sahlgren et al. [17] were pioneers in using the term "Bag-of-Concepts" (BoC) which means that when making the conditions for the term construction (concept), the taxonomy of knowledge is used. BoC then performs conceptualization within a short text in a manner that recognizes the content of the text and then connects the shorter part of the text with the relevant terms-concepts. The aim of conceptualization is an extraction of the most characteristic terms-concepts which best describe the given text [18, 19]. In fact, when in dealing with short texts there is often the risk of losing the context, which makes this procedure useful in the case of working with texts including various types of data, various texts with categorical or topical information, and thus facilitate the work of many applications. This process of mapping short texts within the whole collection has been successfully implemented. This paper aims to map the short text in the context of Law Article and represents it as a vector that best describes a particular part of criminal law when it is not predefined by a legal expert. To reach the representative vectors, one needs to overview the existing methods which lead to the keywords.

Keywords are the ones which are most frequently used in text documents for improving the information retrieval within the sets/collection of documents [20]. Automatic identification of the keywords is a method for analysing words and phrases within small sets/collections so that the content of a given document can be described. The process of identifying keywords aims at accelerating the search of documents in order to obtain the required information. In this regard, four key method extraction categories have been identified: the statistics, linguistics, machine learning and hybrid approach [21]. The statistical approach emphasizes the frequency of the file in a document using four types of frequency points within a set/collection of documents. These are: frequency of files in each subdocument with and without resulting words and frequency in each document with and without resulting words. The text-search techniques primarily establish a series of words and phrases which are then extracted on the basis of statistical analysis of the given factors, such as frequency and distribution. As part of the work process, modern tools create the Term Document Matrix (DTM) and use numerical statistics, so-called TF-IDF [22].

The conventional, TF-IDF weight scheme was invented by Berger et al. [23], which is as follows:

TF: $$TF(t, d_i) = \frac{n_{t,i}}{\sum_{k=1}^{|T|} n_{k,i}}, \quad (3)$$

where $TF(t, d_i)$ is term frequency of word t in document d_i, $n_{t,i}$ is number of occurrences of term t in d_i and $n_{k,i}$ is number of occurrences of all terms in d_i.

IDF: $$IDF_t = \log \frac{M}{m_t + 0.01}, \quad (4)$$

where M is total number of documents in the text corpus and m_t is total number of documents in the text corpus where term t appears.

TF-IDF: $$w(t, d_i) = TF(t, d_i) \times IDF_t, \quad (5)$$

where $w(t, d_i)$ is weight of term t in document d_i.

The frequently used algorithm for indexing TF-IDF requires the existence of a text corpus of documents to use key words of its age. In the absence of a text corpus, keywords can be extracted based on the number of impressions available at the given document. Table 1 shows the seven most often used terms in three separate documents which are found in the studied Articles of the Criminal Code of the Republic of Serbia. It is applied using a morphological normalization vocabulary. The results of the normalization carried out by cutting off the N-gram length of four are presented in Table 2.

Table 1

Normalization 1

No.	TERM	Freq
1.	delo	3
2.	učinilac	3
3.	povreda	2
4.	nastupila	2
5.	zdravlje	2
6.	telesna	1
7.	teška	1

Table 2

Normalization 2

No.	TERM	Freq
1.	delo	3
2.	učin*	3
3.	kazn*	3
4.	zatv*	3
5.	tele*	3
6.	tešk*	3
7.	povr*	3

Shortening the words to 4 characters increases the number of concepts covered key words since, for example, the word “kazniti” and “kaznom” were shortened to

the same base, hence the new key word N-gram “kazn”, as the beginning of the word “kazniće”.

It was established that the use of other algorithms offered in the Optimization Algorithm Toolkit (B-cell Algorithm (BCA), Cloning, Information Gain, Aging (CLIGA), CLONALG, simple and immune algorithm (SIA)), also lead to accurate results although they differ in the duration of operation time.

4 Framework for the Proposed Quick Answers Web Service

In order to create the novel mapping methods, the authors applied the BoC approach (TF-IDF weighting) within the three different documents which are part of the Criminal Code. The candidates are faced with the following combination of the terms: TF values and time IDF values. TF-IDF cannot measure the frequency of keywords in relation to the expected result of a larger set/collection of documents. Search Engine Optimization (SEO) [24], during TF-IDF term weighting in relation to term ranking, performs slightly better than the use of key words. TF-IDF represents each text data as BoW. In the entire text corpus, greater emphasis will be given to the high frequency terms, compared to those, whose frequency is lower. The tendency of TF-IDF is to refine the common terms.

The aim of the Question Answer application was so that citizens can receive quick and accurate answers to their questions. The most common types of content in such an environment are questions, documents and answers. The user may pose a question in various fields. Since the documents are clustered based on keywords, the whole collection is searched in order to find similarities within existing documents from related groups. In the present case study, the citizens' questions refer to criminal offenses as defined in the aforementioned three Articles of criminal law, Articles 121, 122 and 135. Parts of the new law in this QA system are represented as three different document groupings as a whole, comprising the entire currently available knowledge base. To complement and complete this centralized repository, the expert answers were used from the web portal for free legal aid – Pro Bono [25]. The analysis of the existing responses to the questions of citizens concern the criminal acts of the three Articles within the section of Questions and Answers in the field of criminal law. The authors selected 45 representative questions from the given set of questions. These consist of a possible short text documents group that will help towards a better presenting of the considered BoC.

In contrast, finding information from web pages for completing the knowledge base assumes the use of different methods suitable for certain areas that were carried out automatically, which are adaptable and can integrate the data found

[26]. Google's unique and improving algorithm has made it one of the most popular web search engines of all time. For this reason, the authors used Google's search engines to find the content of any tags with keywords from the given BoC. Search engines recognize relevant web pages as additional human resources for a better explanation of the terms that could be important in the mapping of relevant issues with a group of appropriate responses. During the setting up of the topic, the web administrators define and link the texts' dates with a series of key words via the option Related Articles or via the meta keywords tags. In order to discover the group that applies to a given problem, the authors set up a search query for the term "serious bodily injury" as a keyword obtained from electronic dictionaries for the Serbian language via web Language resources exploration site page [27]. The repository, which proved to be the best, was the site of the daily newspapers "Blic". From this site, a total of 35 texts were identified, which corresponded to the specified tag.

4.1 Definition of the Frequency of Term

The minimum term frequency is obtained by the second occurrence of a given term - while in the document this is used to represent the threshold. The term must exist at least twice in a document so as to be counted. Considering the length of the documents to be counted, the following can be stated: supposing that the length of the document is $c_1, c_2, ..., c_n$ and $f(c_1, c_2, ..., c_n)$, representing its frequency, then $c_1, c_2, ..., c_n$ are extracted as terms from the text only if $f(c_1, c_2, ..., c_n)$ is equal to or greater than 2 [28], [29].

The mid interval is the most frequently selected: $(v_i+v_{i+1})/2$ as a representative start. However, C4.5 at the beginning chooses a lower value v_i for each interval $\{v_i+v_{i+1}\}$, not only for the high value [30]. This led the present authors to also implement this type of threshold calculated.

By using stop-words, both the relevant and irrelevant terms within the text were extracted since irrelevant words or concepts often appear. The relevant list of stop-words, for example, can be obtained by moving all the stop-words right from certain words such as the articles (e.g. the, a, an), the preposition (e.g. above, across, before), or some adjectives (e.g. good, nice) in an application [31]. In Serbian, currently there is no formal list of stop-words on the Web site https://sites.google.com/site/kevinbouge/stopwords-lists, thus the authors used the Serbian Lemmatized and PoS Annotated Corpus (SrpLem Kor). The stop-words were created using a list of 700 Serbian stop-words. Moreover, the list of stop-words is automatically constructed in each training process by retrieving the terms with the greatest frequency in the whole Corpus.

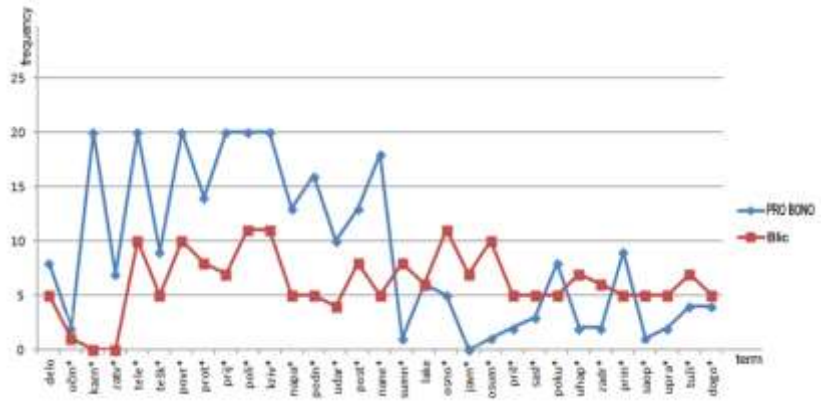


Figure 1
BoC representation in the crime domain for part of the Criminal Code

The TF-IDF weighting scheme is often used in the vector space model (VSM). This model, as well as some distance/similarity measures, are used to determine the similarity between documents or documents and queries represented in the vector space. A document is represented in the t-dimensional term vector space by $d_j = (w_{1,j}, w_{2,j} \ldots w_{t,j})$, where $w_{t,j}$ is defined by the weight key word or terms of the BoC.

In today's web search engines, when the user submits a query to a search system, first they must determine which pages in the index are related to the query and which are not. The proposed web system for quickly answering in the domain of crime, as presented in Figure 2, is based on the absence or presence of it in a document - representation of documents as the BoC. The concepts are mapped to a crime vocabulary for the three law Articles given as vectors below:

- Article 121 [kazn, zatv, tešk, post, javn, poku, tuži, pril, delo, tele, povr, nane, sumn, uhap];
- Article 122 [tele, poli, napa, udar, post, nane, učin, kazn, zatv, kriv, javn, povr, lake, prij, tešk, poku, tuži];
- Article 135 [prij, poli, kriv, post].

One the one hand, this approach is based on the classification of questions on the Criminal Code by using a special domain based on the BoC corpus, while on the other hand, it is based on the comparison of characteristics with the properties of the Codes in the case of the used expert representation of documents.

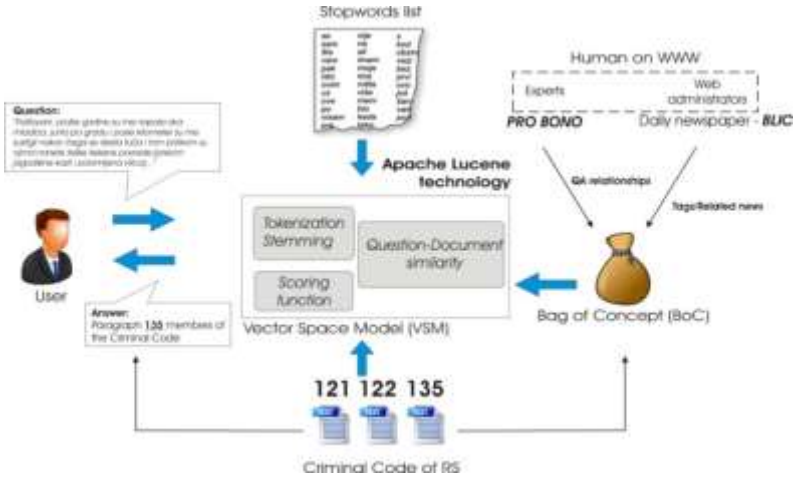


Figure 2

Question Answering system based on the BoC model

In the proposed system, first, it was necessary to map the question to the concepts from the BoC in the corpora defined according to 31 terms seen in Figure 2, presented as a bag. The second step for questions tokenization was to filter the stop-words. Following this, the next step was to remove variable suffixes. This led to a form of morphological normalization (or creating general characteristics). For this reason, when searching texts written in Serbian, the 4-gram analysis was applied. Finally, the authors reached a similar result, starting from posing the question and the BoC display view of three documents using scoring the tool.

The authors' QA system generates as output the relevant document for this question. As a final output, the user then receives the following message as an answer: "Look at Article *n* of the Criminal Code". The value of the number *n*, displayed to the user determines the part of the system software called domain-specific stemming annotator. A stemming annotator is a software agent that uses the BoC source of a particular domain as knowledge bases to map the extracted terms. It assigns the absence or presence of each extracted term in accordance with its keywords relevance within the BoC. Translating the question (query) and documents from raw strings into a computable form is the first hurdle to be overcome in computing a similarity score. To achieve this, the textual representation of information is converted into an algebraic vector space representation [32].

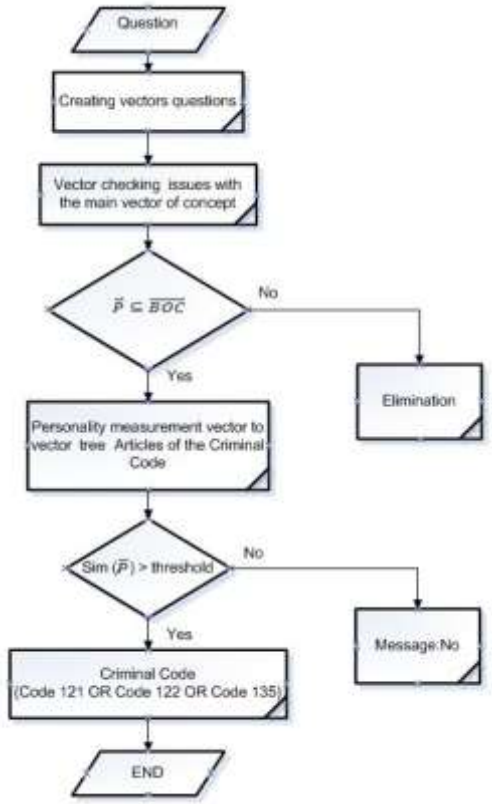


Figure 3
Classification questions algorithm

4.2 Question Classification

In order to ensure the successful operation of the QA system as a whole, it is necessary to perform the correct classification of questions and expected answers. The aim of the classification of the expected questions and answers is mapping - marking the QA types [33]. The targeted values were established on the basis of the values which provided the best results in the previous cases of the text classification. This paper proposes a new algorithm based on the BoC for automating question-document classification tasks such as recommendation on relevant documents. The BoC is a list of all words ranked according to their discriminative power of the three Articles of Criminal Law (clusters). Different measures of distance or similarity use the vectors from each document to calculate distances from the question.

Once a set of representative terms is selected for each cluster $c_j \in C$ $(= \{c_1, c_2, \ldots, c_n\})$, the set of terms is regarded as the representative terms (rt_j) of cluster c_j. This is followed by the comparison of the similarity of each question q_i mapping in the BoC (matching the vector of the question with each vector cluster) with each rt_j using the certain similarity metric to automatically compute a relevance score that measures the similarity between the question and Article number of the Crime Code - c_j. In order to determine the threshold of similarity between the question presented in the form of a short text and the Articles of the law, which is also shown as a short text, the following formula was used [34]:

$$Similarity = \frac{W(Sa) \cap W(Sb)}{\min(W(Sa), W(SB))}, \tag{6}$$

where $W(S_a) \cap W(S_b)$ is an intersection set number of words in questions q_i and the number of words in rt_j, and $min\ (W(S_a),\ W(S_b))$ is a value lower than the number of words, and in both documents.

5 Experimental Evaluation

Before the process of the clusterization, the similarity/distance measure must be determined. Similarity measure is vital due to its direct influence on the ranking of documents, in fact, direct influence on the degree of closeness or separation of the target documents. In addition, measuring the similarity of documents on the basis of characteristics that depend on the type of data in context to documents and observation leads to the clustering and grouping documents by clusters. There is no measure that is universally optimal for all kinds of clustering problems.

The presented system used the free open-source Lucene library search by Apache corporation. This search method can be embedded in the application code that runs in the background when the need arises or it can be run on a remote website. In the process of searching, the field values are searched. In Lucene, starting with the version 5.0, the *SerbianNormalizationFilter* is included, which normalizes Serbian Cyrillic and Latin characters into "bold" Latin. Cyrillic characters are first converted to Latin, then the Latin characters have their diacritics removed with the exception of *đ,* which is converted to *dj*. Note that this process expects lower-cased input.

public class **SerbianNormalizationFilterFactory**
extends **TokenFilterFactory**
implements **MultiTermAwareComponent**
Factory for **SerbianNormalizationFilter**

For the here-described research, the Criminal Code of the Republic of Serbia was used, which was written in Serbian, in Latin script. Unlike English, Serbian contains the following signs: č, ć, ž, đ and š. In order for these signs to be

identified and then the process of indexing and search executed, it is a pre-requisite that Lucene be supported for the Serbian language. This support is provided by the Lucene version 5.2.0, which was implemented in this study.

Choosing an appropriate similarity measure is crucial for cluster analysis, especially for a particular type of clustering algorithms. Four types of the most used similarities were included so that the one be chosen which provides the most precision results for the Criminal field.

The following text similarity functions were used:

1. Cosines Similarity

$$SIM_C(\overrightarrow{t_a}, \overrightarrow{t_b}) = \frac{\overrightarrow{t_a}.\overrightarrow{t_b}}{|\overrightarrow{t_a}|^2 \times |\overrightarrow{t_b}|^2}, \quad (7)$$

where t_a and t_b are m-dimensional vectors over the term set $T = t_1, ..., t_m$.

2. Jaccquard's Coefficient

$$SIM_J(\overrightarrow{t_a}, \overrightarrow{t_b}) = \frac{\overrightarrow{t_a}.\overrightarrow{t_b}}{|\overrightarrow{t_a}|^2 + |\overrightarrow{t_b}|^2 - \overrightarrow{t_a}.\overrightarrow{t_b}} \quad (8)$$

The Jaccard's coefficient is a similarity measure and ranges between 0 and 1. It is 1 when $t_a = t_b$ and 0 when t_a and t_b are disjoint where 1 means that the two objects are identical whereas 0 means they are completely different. The corresponding distance measure is $d_j = 1 - SIM_j$.

3. Euclidean Distance

$$D_E(\overrightarrow{t_a}, \overrightarrow{t_b}) = \left(\sum_{t=1}^{m} |w_{t,a} - w_{t,b}|^2\right)^{\frac{1}{2}}, \quad (9)$$

where the term set is $T = t_1, ..., t_m$. As mentioned previously, the TF –IDF value was used as term weights, that is $w_{t,a} = tf\text{-}idf(d_a, t)$.

In order to determine the appropriate measure of similarity in this proposed system, a study was conducted using the similarity of all three above-mentioned measures over 10 questions in the form of text documents in the field of the Criminal Code. The results of this analysis are presented in Table 3.

Table 3
A summary table of similarities

	Article 121	**Article 122**	**Article 135**	**Expert**
Question KP 1	sim(Cos)= 0.034003 sim(Jacc.)=0.909091 sim(Eucl.)= 17.320508	sim(Cos)= 0.031068 sim(Jacc.)= 0.909091 sim(Eucl.)= 7.937254	sim(Cos)= **0.188982** sim(Jacc.)= **1.000000** sim(Eucl.)= **5.291503**	Article 135
Question KP 2	sim(Cos)= 0.230796 sim(Jacc.)= **1.000000** sim(Eucl.)= 13.152946	sim(Cos)= 0.312195 sim(Jacc.)= 0.909091 sim(Eucl.)= 6.708204	sim(Cos)= **0.435801** sim(Jacc.)= **1.000000** sim(Eucl.)= **8.426150**	Article 121

Question KP 3	sim(Cos)= 0.154287 sim(Jacc.)= 0.916667 sim(Eucl.)= 16.792856	sim(Cos)= 0.197359 sim(Jacc.)= 0.916667 sim(Eucl.)= 7.549834	sim(Cos)= **0.268687** sim(Jacc.)= **1.000000** sim(Eucl.)= **7.000000**	Article 122, 135
Question KP 4	sim(Cos)= **0.522323** sim(Jacc.)= 0.900000 sim(Eucl.)= 15.427249	sim(Cos)= 0.434959 sim(Jacc.)= **1.000000** sim(Eucl.)= **5.477226**	sim(Cos)= 0.268687 sim(Jacc.)= **1.000000** sim(Eucl.)= 7.000000	Article 121, 122
Question KP 5	sim(Cos)= 0.402331 sim(Jacc.)= 0.888889 sim(Eucl.)= 15.905974	sim(Cos)= 0.441129 sim(Jacc.)= 0.888889 sim(Eucl.)= 5.477226	sim(Cos)= **0.759257** sim(Jacc.)= **1.000000** sim(Eucl.)= **2.828427**	Article 135
Question KP 6	sim(Cos)= **0.017314** sim(Jacc.)= 0.909091 sim(Eucl.)= 16.970563	sim(Cos)= 0.000000 sim(Jacc.)= 1.000000 sim(Eucl.)= 7.000000	sim(Cos)= 0.000000 sim(Jacc.)= **1.000000** sim(Eucl.)= **3.464102**	Article 135
Question KP 7	sim(Cos)= 0.660926 sim(Jacc.)= **1.000000** sim(Eucl.)= 15.394804	sim(Cos)= 0.648886 sim(Jacc.)= 0.750000 sim(Eucl.)= 5.099020	sim(Cos)= **0.92582** sim(Jacc.)= **1.000000** sim(Eucl.)= **4.690416**	Article 121

To select the measure, the results obtained with the 3 date rates were compared to the similarities given over the formula (as indicated in the previous section). As in the case of similarity, it is demonstrated that the question Q1 has the greatest similarity with Article 135 where the value was 1.000000. If certain similarities in the column of Article 135 are detected, it can be determined where they are equal or greater in value than is given by the Jaccard similarity. However, the question arises if it was taken as a reference similarity measure. The remaining six issues were analysed in the same manner. The results show that the reference rate of similarity is in this case the Jaccard's correlational coefficient, which is taken as a representative for the calculation of the algorithm in Section 4 in Figure 3.

The results of checking the validity (accuracy) of this algorithm on the basis of the types of alignment by the gold standard vs. real prediction is summarized in the table below:

Table 4

Evaluation Metrics: Classification View

Artical \ Expert	Retrieved	Not Retrieved
Relevant	Relevant Retrieved	Relevant Rejected
Not relevant	Irrelevant Retrieved	Irrelevant Rejected

$$Precision = \frac{Relevant\ Retrieved}{Retrieved} = 75{,}71\% \quad (10)$$

$$Recall = \frac{Relevant\ Retrieved}{Relevant} = 57{,}00\% \quad (11)$$

$$F_{i\ (i=1,n)} = \frac{2*precision_i*recall_i}{precision_i+recall_i} \quad (12)$$

$$F_{Average} = \frac{F_1+F_2+\cdots+F_n}{N} \quad (13)$$

Precision, recall and F_1 are only focus on true positives, i.e., those positive examples by the gold standard. In a monolingual alignment, positive examples are those tokens that are aligned, while negative examples are those that are not. Usually the focus is only on whether those that should have been aligned, are indeed correctly aligned, thus the measure of F_1 is a good *t*.

If correct rejection (true negative) is important, then accuracy is computed instead:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

Accuracy weighs true positives and true negatives equally. The aim is to ensure that the Classifier recognizes both the positive and negative examples. Specifically, in alignment, where most tokens are not aligned, the accuracy value is likely to be very high in general and it is difficult to determine the difference. In such an instance, only F_1 on positive (aligned) examples is reported.

Number of questions extractions results- In order to test the algorithm, the authors used 10 questions. Table 5 shows the results obtained on the basis of the above-presented algorithm, further, the table also displays the results provided by the expert in the field of the Criminal Code.

Table 5
Table of results

Precision	75,71 %
Recall	57,00 %
$F_{Average}$	0.6936
TP	7
FP	5
FN	3
Accuracy	46,66 %

Conclusions

This paper deals with the impact of certain parameters of algorithms on the time period required for processing, trying to determine the optimal value of certain algorithms. The examples analysed in this paper present unique algorithms resulting in optimal solutions using different time values in relation to different parameter values. In this regard, multiple optimal values exist as extreme, unacceptable values.

In the authors' view, it is also important to check and research the behaviour of each of these algorithms in relation to the number of different parameters, which comprises the practical part of this paper. It has been confirmed that all available algorithms are not applicable in certain situations (function optimization), given that specific algorithms analysed during the study did not recognize certain extreme test search functions.

A new model, based on the BoC approach in text representations, was presented in the paper. The proposed system was focused on the terms and they failed to identify concepts when attempting to understand the text. Using the BoC is suggested in order to improve the performance of QA applications. This approach was achieved by the novel exploration of a set of predefined classes and has provided impressive results in the process dimensionality reduction by the mapping query to the document. The research shows that this current system can map queries to the relevant code Article with average precision values of 75,71 %. In general, citizen queries in e-Government services must be very short and they are difficult to classify using a traditional classifier technique, such as the BoW. The documents or parts of them need to be processed so as to obtain typical concepts since the queries can be similar in the main topics of the documents. Thus, much time is saved by not having to process large amounts of long text so as to enable the concept model of documents presented as predefined classes to give quick recommenders presented as answers.

Regarding the concepts as sub contents, these can diversify the recommendations directly from the subtopic level where a large knowledge base is needed to convert the terms to concepts. Helping to adequately construct a concept model for each member of the law requires the use of many sources. Therefore the authors recommend the use of existing news Articles by people already grouped and classified using the tag of that mark. This principle of addition to the relevant terms of the existing unstructured documents is necessary to describe what precise vector the document is given in the absence of a dictionary for the area. This framework is suitable for many future e-services, especially those that do not contain pre-set response answers offered in the form of relevant documents. Also, the authors found that using the Jaccard index provided the best results in comparison to other similarity measures for comparing short text, which might be suitable for short text such as the clustering process short-text corpora. This paper presents one solution of an e-government service application designed to find the existing answers or relevant document for citizens' questions asked in Serbian. The research presented in this paper indicates that a new approach must be introduced with the aim to solve similar problems in current and future Serbian public services.

Acknowledgement

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia under the project no. III 47016.

References

[1] G. Šimić, Z. Jeremić, E. Kajan, D. Randjelović, A. Presnall, A Framework for Delivering e-Government Support, *Acta Polytechnica Hungarica*, 11(1), 2014, pp. 79-96

[2] A. Medelyan, NLP keyword extraction tutorial with RAKE and Maui, https://www.airpair.com/nlp/keyword-extraction-tutorial. [Accessed April 2016]

[3] D. Avdic, A. Avdic, Spalević, U. Marovac, M-Government Application Intended to Search Documents Written in Serbian Language, in *Sinteza 2014 - Impact of the Internet on Business Activities in Serbia and Worldwide*, Belgrade, Singidunum University, Serbia, 2014, pp. 902-906, doi: 10.15308/sinteza-2014-902-906

[4] U. Marovac, A. Pljasković, A. Crnišanin, E. Kajan, N-gram analysis of text documents in Serbian, Proceedings of 20th Telecommunications forum TELFOR 2012, pp. 1385-1388

[5] R. Stanković, C. Krstev, I. Obradović, O. Kitanović, Indexing of Textual Databases Based on Lexical Resources: A Case Study for Serbian, Semantic Keyword-Based Search on Structured Data Sources, *Springer*, 2016, pp. 167-181

[6] O. Kolomiyets, M.F. Moens, A survey on question answering technology from an information retrieval perspective, *Information Sciences* 181(24), 2011, pp. 5412-5434

[7] L. Wenyin, T. Hao, W. Chen, M. Feng, AWeb-Based Platform for User-Interactive Question-Answering, *World Wide Web* 12(2), 2009, doi 10.1007/s11280-008-0051-3, pp. 107-124

[8] "The Nature of Mathematical Programming", Mathematical Programming Glossary, *INFORMS Computing Society*, https://glossary.informs.org/ver2/mpgwiki/index.php?title=Extra:Mathematical_programming [Accessed 15 March 2017]

[9] T. W. Miller, Data and Text Mining: A Business Applications Approach, *Pearson / Prentice Hall*, New Jersey, 2005

[10] G. Koteswara Rao, D. Shubhamoy, Decision Support for E-Governance: A Text Mining Approach, *International Journal of Managing Information Technology (IJMIT)*, 3(3), August 2011, DOI: 10.5121/ijmit.2011.3307. pp. 73-91

[11] Y. Yang, T. Joachims, Text Categorization, Scholarpedia, 3(5): 4242. http://www.scholarpedia.org/article/Text_categorization. doi:10.4249/scholarpedia.4242. [Accessed 11 March 2017]

[12] W. Jiang, B. Samanthula, N-Gram based Secure Similar Document Detection, the 25th Annual WG 11.3 Conference on Data and Applications Security, Richmond, Virginia, July 11-13, 2011

[13] R. Kumar, V. Mansotra, Applications of Stemming Algorithms in Information Retrieval - A Review, *International Journal of Advanced*

Research in Computer Science and Software Engineering 6(2), February - 2016, pp. 418-423

[14] D. Subotić, N. Forbes, "Serbo-Croatian language – Grammar", *Oxford Clarendon press*, pp. 25-31, 61-64, 101-113

[15] Paragraf - legal and economic editions, www.paragraf.rs. [Accessed 1 Jun 2017]

[16] F. Wang, Z. Wang, Z. Li, J-R Wen, Concept-based Short Text Classification and Ranking, *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2014. pp. 1069-1078

[17] M. Sahlgren, R. Cöster, Using bag-of-concepts to improve the performance of support vector machines in text categorization, In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING 2004, pp. 487– 493, Geneva, Switzerland, August 2004

[18] Y. Song, H. Wang, Z. Wang, H. Li, W. Chen, Short text conceptualization using a probabilistic knowledgebase, *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three*, AAAI Press, 2011, pp. 2330-2336

[19] Z. Wang, H. Wang, Z. Hu, Head, modifier, and constraint detection in short texts, In: *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*. IEEE, 2014. pp. 280-291

[20] A. Hulth, Improved automatic keyword extraction given more linguistic knowledge, *Proceedings of the 2003 Conference on Emprical Methods in Natural Language Processing*, Sapporo, Japan, 2003, pp. 216-223

[21] C. Zhang, H. Wang, Y. Liu, D. Wu, Y. Liao, B. Wang, Automatic Keyword Extraction from Documents Using Conditional Random Fields, *Journal of Computational Information Systems*, 4(3), 2008. pp. 1169-1180

[22] G. Wunnava, Applying Machine Learning to Text Mining with Amazon S3 and RapidMiner (on Jun 25 2015), AWS Big Data Blog. https://blogs.aws.amazon.com/bigdata/post/Tx22THFQ9MI86F9/Applying-Machine-Learning-to-Text-Mining-with-Amazon-S3-and-RapidMiner. [Accessed March 2017]

[23] A. Berger et al, Bridging the Lexical Chasm: Statistical Approaches to Answer Finding, In: *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2000. pp. 192-199

[24] A. B. King, Website Optimization, *O'Reilly Media*, July 2008

[25] Pro Bono - portal for free legal aid, http://www.besplatnapravnapomoc.rs/

[26] I. Augenstein, D. Maynard, F. Ciravegna, Distantly Supervised Web Relation Extraction for Knowledge Base Population, *Semantic Web* 7(4), 2016. pp. 335-349

[27] Human Language Tecnologies, http://hlt.rgf.bg.ac.rs/VeBranka/BagOfWords.aspx . [Accessed 3 May 2017]

[28] S. Li, H. Wang, S. Yu, C. Xin, News-Oriented Keyword Indexing with Maximum Entropy Principle, *Proceedings of the 17th Pacific Asia Conference on Language, Information and Computation*. 2003. pp. 277-281

[29] S. Li, H. Wang, S. Yu, C. Xin, News-Oriented Automatic Chinese Keyword Indexing, *Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17*. Association for Computational Linguistics, 2003. pp. 92-97

[30] J. R. Quinlan, C4.5: Programs for Machine Learning, *Morgan Kaufmann Publishers Inc.*, San Francisco, CA, USA, 1993

[31] K. Ganesan, All About Stop Words for Text Mining and Information Retrieval (on Oct 6, 2014), http://www.text-analytics101.com/2014/10/all-about-stop-words-for-text-mining.html#sthash.V4f3jaL1.dpuf. [Accessed 17 Jun 2017]

[32] T. Magerman et al., Assessment of Latent Semantic Analysis (LSA) text mining algorithms for large scale mapping of patent and scientific publication documents, 2011

[33] D. Metzler, W. B. Croft, Analysis of Statistical Question Classification for Fact-based Questions, *Information Retrieval*, 8(3), 2005.pp. 481-504

[34] S. S. Singh, Statistical Measure to Compute the Similarity between Answers in Online Question Answering Portals, *International Journal of Computer Applications* (0975 – 8887) Volume 103 (15), October 2014

Modeling Synchronization Problems: From Composed Petri Nets to Provable Linear Sequents

Ján Perháč, Daniel Mihályi, Valerie Novitzká

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9 042 00 Košice, Slovakia
Jan.Perhac@tuke.sk, Daniel.Mihalyi@tuke.sk, Valerie.Novitzka@tuke.sk

Abstract: Component-based programming has become a popular and frequently used method for software development, prepared from independent components by a composition. Our paper presents an illustration of how a composition of components can cause the emergence of new problems that should be solved, in order to obtain the desired results. We introduce a transformation of Petri nets to corresponding provable sequents of linear logic and then we show how a composition of simple Petri nets causes arising of synchronization problems. Transformation to provable linear sequents that ensures a verified behavior for such composed systems.

Keywords: linear logic; petri nets; component composition

1 Introduction

In the last decades, a new method of software development, a component-based programming, has become very popular. It can be characterized as a construction of program systems from working entities by their composition. There are many publications in the area of software engineering discussing basic principles of this method and providing practical guides for how to prepare such program systems, e.g. [21, 22, 23, 27, 29]. Among the basic properties of components is their independency, i.e. that they can be executed separately, can be reused, and should satisfy rules of composition (contracts and dependencies) to ensure that the whole system produces desired results. One of the important properties of a composition is that after it is composed, new problems can arise. The aim of our paper is to illustrate this situation on known examples of synchronization. We used two formal tools; Petri nets (PNs) and linear logic. Both tools are useful and frequently employed in modeling and behavioral description of program systems. We show how the behavior of PNs can be expressed by provable sequents of linear logic, and simultaneously, we show how the synchronization problems arise from the composition of simple components.

Petri nets are well-known mathematical and graphical tool for behavioral modeling of processes. They were designed especially for modeling systems with interacting concurrent processes [20, 30]. They are sometimes called place/transition nets and they enable graphical expression of the system execution. Petri nets nowadays have a wide spectrum of applications, not only in concurrent programming, but also in business process modeling, data analysis, software design, simulation, and many others [2, 3, 9, 10]. The expressive power of Petri nets has been studied and compared with several other formal methods for the specification and description of systems. Such systems include algebraic specification in [24], which leads to the extension of the notion of tokens to structural ones, or with B-language [11], in which the transformation of Petri nets to B-language is defined for software development, from specification to implementation. In our paper we would like to present another correspondence, namely how PNs can be transform into provable linear logic sequents.

Linear logic is a logical system introduced by Jean-Yves Girard in 1987 [6, 7]. It enables description of processes as they behave in real world. It is capable of describing the dynamic of processes, parallelism, external and internal nondeterminism, consecutive processes, and it is also able to handle the resources on syntactic level. Thanks to these properties, linear logic can be considered as a bridge between computing science and logic [15]. Propositional linear logic is available for program systems description [1, 16], their behavior [26], and their extension with modal operators enables modeling of knowledge achievement [17, 18] mainly in intrusion detection systems to improve network reliability [28]. Linear logic has greater expressive power than classical logic, thanks to more connectives with special properties. Moreover, every formula of classical propositional logic can be expressed in linear logic. One of the most important properties of linear logic is its ability to describe the dynamic of processes. By linear implication $\multimap$ it is possible to express sequentiality and causality of processes.

Another important property of linear logic is its ability to handle resources - logical space and logical time [14, 19]. It allows us to express the internal structure of the resources, their consumption, together with a continuation of processes in incremental time.

In our research, we were inspired by the approach published in [5], where PNs are defined as models of linear logic. The reverse property is in [13]. We follow up on this approach in a way by which each significant fragment of a PN can be described by a linear sequent. We formulate a transformation of Petri nets to provable linear sequents. To illustrate composition problems, we define a trivial working PN and show its behavior by corresponding provable linear sequent. Composing two and five trivial PNs causes new problems to arise, namely synchronization problems of mutual exclusion, and deadlock. These problems are similar to the ones in the process of component composition in component-based systems. The examples of synchronization problems appear frequently in

textbooks or papers concerning with Petri nets [9, 11, 20, 25]. To illustrate the transformation from PNs to linear logic together with the appearance of synchronization problems in component composition, is our main aim that is not published yet in scientific publications.

2 Principles of Petri Nets

A Petri net is a known formal tool used mainly for modeling the behavior of concurrent systems as state-transition systems. Their advantage is that these models can be represented graphically, by directed bipartite graphs. A PN graph has two types of nodes: places and transitions. *Places* are represented by circles and *transitions* by arrows (arcs). Places express possible states of a system, and transitions represent changes of states, i.e. events. Places can contain special marks called *tokens.* A transition is enabled, i.e. it can be *fired*, if all places of input arcs contain required number of tokens. When a transition is fired, it produces tokens in all places on output arcs. Generally, execution of a PN is nondeterministic: when more than one transition is enabled, any of them can be fired. Any distribution of tokens over places represents a configuration of a given PN called *marking*. For any place p of a PN its marking is a function $m: P \rightarrow \mathbb{N}_0$ returning a number of tokens in p. $\mathbb{N}_0$ denotes the set of natural numbers with zero. A marking of a PN is defined as a tuple

$$m = (m(p_1), \dots, m(p_n)) \tag{1}$$

of markings of all places in a PN.

A PN has two functions: *pre* and *post*. The pre function for each place and transition returns a number of arcs from a given place to a given transition. The $post$ function for each place and transition returns a number of arcs outgoing from a given transition to a given place. Formally, a Petri net is defined as a tuple

$$PN = (P, T, pre, post, m_0), \tag{2}$$

where $P = \{p_1, \dots, p_n\}$ is a finite set of places, $T = \{t_1, \dots, t_n\}$ is a finite set of transitions, $pre: P \times T \rightarrow \mathbb{N}_0$ and $post: T \times P \rightarrow \mathbb{N}_0$ are pre- resp. post-conditions and m_0 is initial marking before a PN starts its execution.

The execution of a PN is based on following two rules:

- *enabling rule* formulates the condition under which a transition is allowed to be fired. A transition can be fired if each of its input places contains a number of tokens greater or equal than the given threshold
- *firing rule* defines modification of a marking caused by firing a transition. When a transition t is fired, a token from each input place is deleted, and to each output place a token is added

A *behavior* of a PN can be described as a sequence of markings reached during the execution of PN.

3 Basic Concepts of Linear Logic

Linear logic belongs to the new logical systems that are useful for describing and verifying of real systems. It facilitates the formulation of dynamic processes, nondeterminism, concurrency, and handling of resources such as time and space on syntactic level [12]. In this section we introduce the basic definitions of logical connectives and the deduction system of linear logic. Elementary propositions of linear logic can be considered as actions or resources.

Let $Prop = \{p_1, p_2, \dots\}$ be a countable set of elementary propositions. In this paper we use the following syntax of linear formulas:

$$\varphi ::= p \mid \mathbf{1} \mid \mathbf{0} \mid \top \mid \varphi \otimes \psi \mid \varphi \oplus \psi \mid \varphi \,\&\, \psi \mid \varphi \multimap \psi \mid \varphi^{\perp} \tag{3}$$

where

- $\varphi \otimes \varphi$ is the *multiplicative conjunction* with neutral element $\mathbf{1}$. This formula expresses that both actions perform simultaneously or both resources are available at once.
- $\varphi \,\&\, \psi$ is the *additive conjunction* with neutral element $\top$. This formula expresses that only one of the actions performs, but we can deduce which one from the environment. Additive conjunction describes external nondeterminism.
- $\varphi \oplus \varphi$ is the *additive disjunction* with neutral element $\mathbf{0}$. This formula expresses that only one of the actions performs but we cannot anticipate which one. Additive disjunction describes internal nondeterminism.
- $\varphi^{\perp}$ is the *linear negation* and it describes a reaction of an action φ or a consumption of a resource φ. Linear negation has the property of involution, i.e. $(\varphi^{\perp})^{\perp} \equiv \varphi$.
- $\varphi \multimap \psi$ is the *linear implication*. This formula describes that the first action φ is a cause of the second (re)action ψ or in the case of resources it expresses that the first resource is consumed after linear implication.

For linear formulas, we use *sequent calculus*. A sequent has the form

$$\Gamma \vdash \varphi, \tag{4}$$

that expresses that a formula φ is deducible from the formulas in the context $\Gamma = \varphi_1, \dots, \varphi_n$. The formulas on the left side of a sequent are assumptions, therefore we consider them as multiplicative conjunction $\otimes$. The formula φ on the right side is deducible from the assumptions.

The sequent deduction system consists of basic rules, rules for the connectives, rules for neutral elements and structural rules:

1. Basic rule:

$$\frac{}{\varphi \vdash \varphi}\ (id)$$

2. Rule for linear negation:

$$\varphi^{\perp} \equiv \varphi \multimap \perp\ ((\,)^{\perp})$$

3. Rules for neutral elements:

$$\frac{}{\Gamma \vdash \top}\ (\top_{-r}) \qquad \frac{}{\Gamma, 0 \vdash \varphi}\ (0_{-l}) \qquad \frac{\Gamma \vdash \varphi}{\Gamma, 1 \vdash \varphi}\ (1_{-l}) \qquad \frac{}{\vdash 1}\ (1_{-r})$$

4. Rules for connectives:

$$\frac{\Gamma_1 \vdash \varphi_1 \qquad \Gamma_2, \varphi_2 \vdash \psi}{\Gamma_1, \varphi_1 \multimap \varphi_2, \Gamma_2 \vdash \psi}\ (\multimap_{-l}) \qquad \frac{\Gamma, \varphi_1 \vdash \varphi_2}{\Gamma \vdash \varphi_1 \multimap \varphi_2}\ (\multimap_{-r})$$

$$\frac{\Gamma, \varphi_1, \varphi_2 \vdash \psi}{\Gamma, \varphi_1 \otimes \varphi_2 \vdash \psi}\ (\otimes_{-l}) \qquad \frac{\Gamma_1 \vdash \varphi_1 \qquad \Gamma_2 \vdash \varphi_2}{\Gamma_1, \Gamma_2 \vdash \varphi_1 \otimes \varphi_2}\ (\otimes_{-r})$$

$$\frac{\Gamma, \varphi_1 \vdash \psi}{\Gamma, \varphi_1 \,\&\, \varphi_2 \vdash \psi}\ (\&_{-l1}) \qquad \frac{\Gamma, \varphi_2 \vdash \psi}{\Gamma, \varphi_1 \,\&\, \varphi_2 \vdash \psi}\ (\&_{-l2})$$

$$\frac{\Gamma, \varphi_1 \vdash \psi \qquad \Gamma, \varphi_2 \vdash \psi}{\Gamma, \varphi_1 \,\&\, \varphi_2 \vdash \psi}\ (\oplus_{-l}) \qquad \frac{\Gamma \vdash \varphi_1 \qquad \Gamma \vdash \varphi_2}{\Gamma \vdash \varphi_1 \,\&\, \varphi_2}\ (\&_{-r})$$

$$\frac{\Gamma, \vdash \varphi_1}{\Gamma \vdash \varphi_1 \otimes \varphi_2}\ (\otimes_{-r1}) \qquad \frac{\Gamma \vdash \varphi_2}{\Gamma \vdash \varphi_1 \otimes \varphi_2}\ (\otimes_{-r2})$$

5. Structural rules:

$$\frac{\Gamma, \varphi_1, \varphi_2 \vdash \psi}{\Gamma, \varphi_2, \varphi_1 \vdash \psi}\ (exch) \qquad \frac{\Gamma_1 \vdash \varphi \qquad \Gamma_2 \vdash \varphi}{\Gamma_1, \Gamma_2 \vdash \psi}\ (\otimes_{-r})$$

The basic rule is the axiom of identity, linear negation is expressed as linear implication. The rules of connectives introduce connectives on the left or on the right part of a sequent. The only structural rules are exchange and cut rules. We note that in linear logic it is important which resources and how many of them we have. Therefore, the obvious structural rules of weakening and contraction are not allowed, only in controlled way using special unary connectives, exponentials. In this paper, we do not use exponentials; their definition together with

corresponding deduction rules is in [8]. A proof in sequent calculus is a tree, where the root is a proven sequent, and every step is the application of an appropriate deduction rule. A sequent is provable if all of the leaves of its proof tree are identities.

4 Transformation of Petri Net Patterns to Linear Logic Sequents

In this section we define transformations of some parts, the patterns of PNs, to the corresponding linear formulas. We select several significant patterns that occur in synchronization PNs and we define corresponding linear logic sequents. All other patterns can be transformed similarly using this idea.

We introduce the following notation. A place p of a PN containing one token, i.e. with the marking $m(p) = 1$, we denote by the elementary proposition p of linear logic. Marking expressing that a place p_1 contains one token and a place p_2 contains two tokens, can be denoted using multiplicative conjunction:

$$p_1 \otimes p_2 \otimes p_2 \tag{5}$$

For describing of a PN transition by linear formula we use linear implication —o. The premise of the implication is a marking that makes a transition t to be enabled and the conclusion is the marking that arises after firing of t. Linear implication expresses change of a state, caused by firing a transition together with the consumed resources (tokens) on the left side and the produced resources (tokens) on the right side of implication [4]. For instance, if t is a transition enabled when the places p_1 and p_2 have both one token and after firing t the place p_3 obtains a token, then this transition can be denoted by the following linear implication:

$$t \equiv p_1 \otimes p_2 \text{ —o } p_3 \tag{6}$$

A behavior of a PN is described by the sequents of linear logic in the form:

$$m, l \vdash m', \tag{7}$$

where

- m is a marking before firing a transition
- l is a list of enabled transitions expressed by linear implications defined above
- m' is a marking after firing a transition

Such sequents express that the marking m' is produced from a marking m by firing a transition from l.

Now, we consider the characteristic fragments of PNs representing their possible structure and we formulate the corresponding linear formulas and sequents.

Precisely:

- places of a PN are transformed into linear formulas
- transitions of a PN are transformed into linear implications

We choose a *sequence* illustrated in Figure 1 as the first pattern. The transition t is enabled if the place p_1 contains at least one token. After firing t, the place p_2 obtains one token.

Figure 1
Sequence

This pattern can be transformed to the following sequent:

$$p_1, t \vdash p_2, \tag{8}$$

where $t \equiv p_1$—o p_2. We call this sequent *causality*, because p_1 is a cause of p_2.

The pattern in Figure 2 depicts a situation of *furcation*:

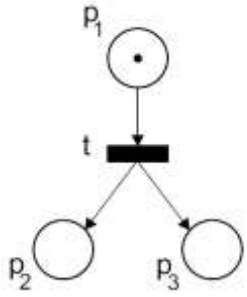


Figure 2
Furcation

This pattern has one transition t that is enabled if at least one token is in p_1. Firing of t produces one token in both p_2 and p_3, simultaneously. Therefore, the corresponding sequent expresses *concurrency*:

$$p_1, t \vdash p_2 \otimes p_3 \tag{9}$$

The pattern in Figure 3 illustrates a situation of *rendezvous*. To be t enabled the places p_1 and p_2 have to have a token and after firing t one token is produced in p_3.

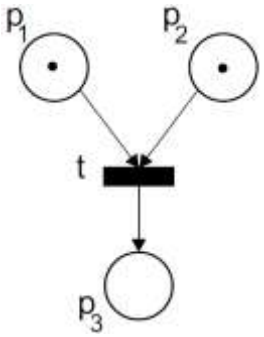


Figure 3
Rendezvous

We transform this pattern to the following sequent expressing *synchronization*:

$$p_1, p_2, t \vdash p_3 \tag{10}$$

Up to now, all patterns were deterministic because they have only one enabled transition.

The next two patterns express nondeterministic behavior. The first of them in Figure 4 is *free choice*. Either t_1 or t_2 are enabled but we cannot decide which of them.

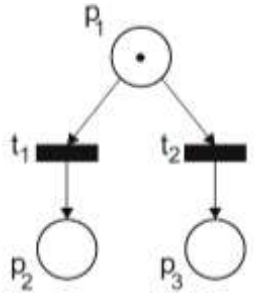


Figure 4
Free choice

We transform this pattern to the sequent using additive disjunction $\oplus$ between formulas describing enabled transitions on the left side, because only one of t_1 or t_2 can be fired, but we do not know which one. It expresses *internal nondeterminism*. Producing tokens on the right side is expressed by linear formula by additive conjunction because a token can be obtained either in p_2 or in p_3 depending on previous action:

$$p_1, t_1 \oplus t_2 \vdash p_2 \& p_3. \tag{11}$$

The last pattern that we consider is *dependent choice* in Figure 5. Dependent or environmental choice expresses the situation when only one transition of t_1 or t_2 can be fired, but it depends on the occurence of a token either in p_1 or in p_2. In the other words, this choice depends on the given environment. Firing any of transition t_1 or t_2 produces one token in the place p_3.

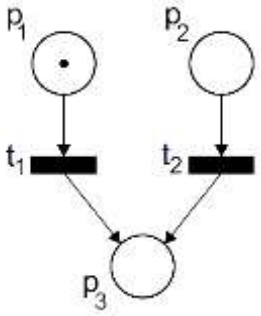


Figure 5
Dependent choice

We transform this pattern to the linear sequent with additive conjunction describing transitions on the left side, because firing of t_1 or t_2 depends on the situation in p_1 or p_2, respectively:

$$p_1 \,\&\, p_2, t_1 \,\&\, t_2 \vdash p_3 \tag{12}$$

5 Synchronization Problems Arising from PNs Composition and Verifying by Proofs

In this section, we present the main aim of our paper. First, we define a trivial PN that works, i.e. its transitions can be enabled and fired. Considering a trivial PN as a component, when we compose two such PNs a new problem arises, that is known as *mutual exclusion*. We transform such composed PN into linear sequent and verify it by a proof. Second, we compose five trivial PNs that lead to arising of a known case of dining philosophers, i.e. deadlock problem. Again, we transform this composed PN into linear sequent and verify it by a proof. To summarize, we illustrate not only transformation of PNs to corresponding linear sequents, but also that composition of components together can cause arising of new problems that have to be solved in order to reach desired system behavior.

We start with a trivial PN in Figure 6. It is clear that this PN works and its behavior is deterministic. With a little imagination, we can call this PN, a *dining philosopher.*

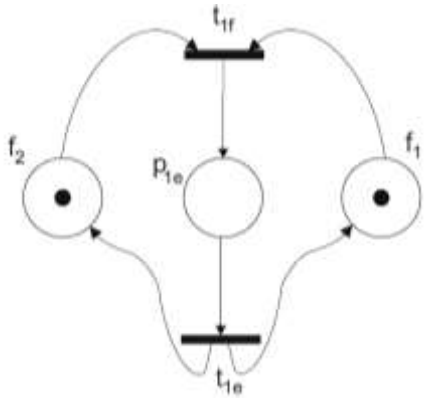


Figure 6
A dining philosopher

This trivial PN consists of three places, where f_1 and f_2 represent forks and p_{1e} represents an eating philosopher. This PN has two transitions t_{1e} and t_{1f} and its initial marking is $m_0 = (1,0,1)$. If both forks f_1 and f_2 are available, i.e. both corresponding places have a token, the transition t_{1f} is enabled and after firing, the place p_{1e} obtains a token, i.e. the philosopher is eating. In this state the transition t_{1f} is enabled and after firing, the philosopher releases the forks and starts to think.

The behavior of this PN can be expressed by the linear sequent:

$$f_2, f_1, (f_2 \otimes f_1) \multimap p_{1e}, p_{1e} \multimap (f_2 \otimes f_1) \vdash f_2 \otimes f_1 \tag{13}$$

with the following proof in the Figure 7:

$$\dfrac{\dfrac{\dfrac{}{f_2 \vdash f_2}\,(\mathrm{id}) \quad \dfrac{}{f_1 \vdash f_1}\,(\mathrm{id})}{f_2, f_1 \vdash f_2 \otimes f_1}\,(\otimes_{-\mathrm{r}}) \quad \dfrac{}{p_{1e} \vdash p_{1e}}\,(\otimes_{-\mathrm{r}}) \quad \dfrac{\dfrac{\dfrac{}{f_2 \vdash f_2}\,(\mathrm{id}) \quad \dfrac{}{f_1 \vdash f_1}\,(\mathrm{id})}{f_2, f_1 \vdash f_2 \otimes f_1}\,(\otimes_{-\mathrm{r}})}{f_2 \otimes f_1 \vdash f_2 \otimes f_1}\,(\otimes_{-1})}{p_{1e}, p_{1e} \multimap (f_2 \otimes f_1) \vdash f_2 \otimes f_1}\,(\multimap_{-1})}{f_2, f_1, (f_2 \otimes f_1) \multimap p_{1e}, p_{1e} \multimap (f_2 \otimes f_1) \vdash f_2 \otimes f_1}\,(\multimap_{-1})$$

Figure 7
Proof of dining philosopher

Now we consider the previous PN in Figure 6 as a component. By composing two such components (Fig. 8), we get the well-known problem of *mutual exclusion* (mutex). The interaction between the components is represented by the place f_2. The principle of mutex is that only one of the processes can be executed in one moment, in other words, if one philosopher is eating (i.e. he possesses his left and right forks), the second is thinking and vice versa. Let the initial marking of this PN be $m_0 = (1,0,1,0,1)$.

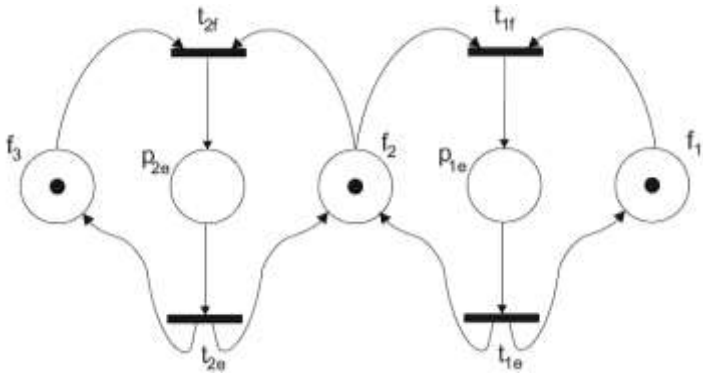


Figure 8
Mutual exclusion

There are two possibilities how this PN works:

- either t_{2f} is fired, which corresponds to the linear logic sequent:
$$f_3, f_2, f_1, (f_3 \otimes f_2) \multimap p_{2e} \vdash p_{2e} \otimes f_1 \quad (14)$$
- or t_{1f} is fired, this case is expressed by linear logic sequent:
$$f_3, f_2, f_1, (f_2 \otimes f_1) \multimap p_{1e} \vdash p_{1e} \otimes f_3 \quad (15)$$

The behavior of mutex can be described by the following linear logic sequent:

$$f_3, f_2, f_1, \big((f_2 \otimes f_1) \multimap p_{1e}\big) \oplus ((f_3 \otimes f_2) \multimap p_{2e}) \vdash (p_{1e} \otimes f_3) \& (p_{2e} \otimes f_1),$$

where we use internal nondeterminism on the left side, i.e. additive disjunction $\oplus$ between transitions t_{2f} and t_{1f}. On the right side of this sequent, we use additive conjunction & between tokens, because they depend on which of the transitions t_{2f} and t_{1f} was actually fired. This sequent is provable and we present a left branch of this proof tree depicted in the Figure 9.

$$\dfrac{\dfrac{\dfrac{\dfrac{}{f_1 \vdash f_1}\,(\mathrm{id}) \qquad \dfrac{\dfrac{\dfrac{}{f_2 \vdash f_2}\,(\mathrm{id}) \quad \dfrac{}{f_3 \vdash f_3}\,(\mathrm{id})}{f_2, f_3 \vdash f_2 \otimes f_3}\,(\otimes_r) \quad \dfrac{}{p_{2e} \vdash p_{2e}}\,(\mathrm{id})}{f_2, f_3, (f_2 \otimes f_3) \multimap p_{2e} \vdash p_{2e}}\,(\multimap_l)}{f_1, f_2, f_3, (f_3 \otimes f_2) \multimap p_{2e} \vdash (f_1 \otimes p_{2e})}\,(\otimes_r) \quad \cdots}{f_1, f_2, f_3, (f_3 \otimes f_2) \multimap p_{2e} \vdash (f_1 \otimes p_{2e}) \& (f_3 \oplus p_{1e})}\,(\&_r) \quad \cdots}{f_1, f_2, f_3, ((f_3 \otimes f_2) \multimap p_{2e}) \oplus ((f_1 \otimes f_2) \multimap p_{1e}) \vdash (f_1 \otimes p_{2e}) \& (f_3 \otimes p_{1e})}$$

Figure 9
Proof of mutual exclusion

The right branch of the proof tree can be constructed similarly. Provability of this sequent ensures that we have a solution of mutual exclusion.

Now, we compose five trivial PNs (components) together and here a new problem arises, known as the problem of *five dining philosophers*, i.e. *deadlock problem* (Figure 10).

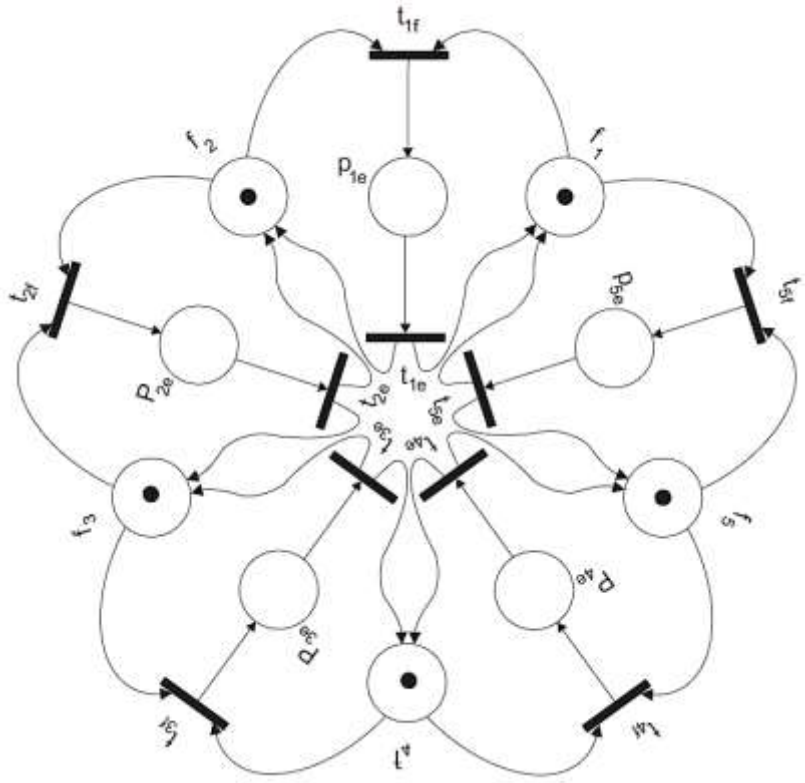


Figure 10
Problem of five dining philosophers

If we assume the order of places $p_{1e}, p_{2e}, p_{3e}, p_{4e}, p_{5e}, f_1, f_2, f_3, f_4, f_5$, then the initial marking is $m_0 = (0, 0, 0, 0, 0, 1, 1, 1, 1, 1)$. If a place p_i, $i = 1, \ldots, 5$ has a token, the philosopher i is eating, if it is empty, he is thinking. The places f_1, f_2, f_3, f_4, f_5 serve for forks, the places $p_{1e}, p_{2e}, p_{3e}, p_{4e}, p_{5e}$ mean that corresponding philosophers eat.

This system works if every philosopher can eat, i.e. each process in a system can be executed. The problem occurs, when each philosopher takes its right fork and then they are waiting forever for the second fork; or in other words, if each process needs a certain resource to be executed, but any of them cannot release resources before finishing their execution. This problem is known as *deadlock*.

From the previous ideas it is clear that either one philosopher can eat, or two not neighbor philosophers can eat at one moment. We describe one of the possible solutions of this problem.

Consider that in the first step one philosopher, e.g. p_{1e} is eating. That means, he takes both forks f_1 and f_2 and the transition t_{1f} is fired. Figure 11 illustrates the system after firing t_{1f}.

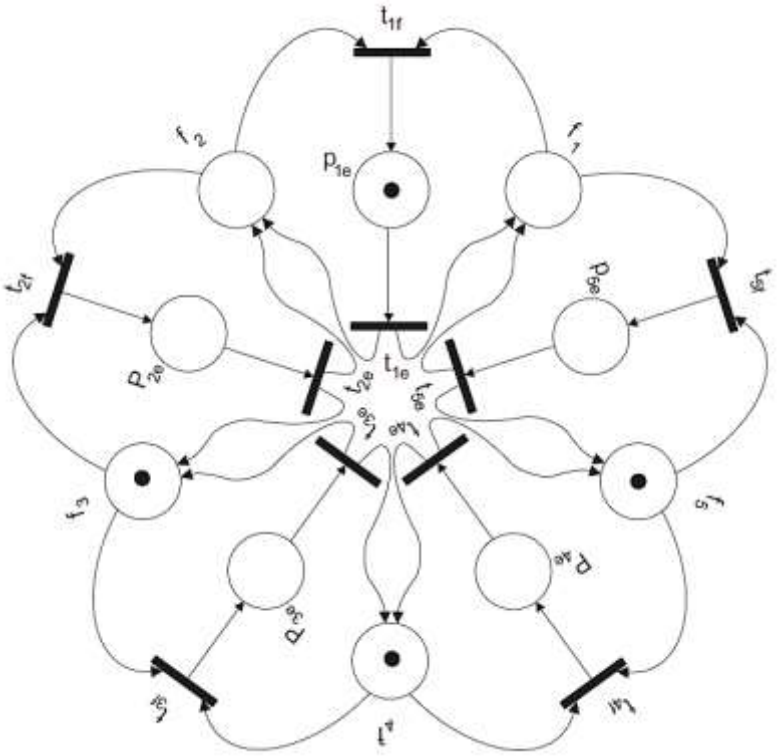


Figure 11
Single dining philosopher

We describe this action by the linear sequent

$$f_1, f_2, f_3, f_4, f_5, f_1 \otimes f_2 \multimap p_{1e} \vdash f_5 \otimes (f_4 \otimes (p_{1e} \otimes f_3)) \tag{16}$$

together with its (fragment of) proof depicted in the Figure 12:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{f_1 \vdash f_1}\,(\mathrm{id}) \quad \dfrac{}{f_2 \vdash f_2}\,(\mathrm{id})}{f_1, f_2 \vdash f_1 \otimes f_2}\,(\otimes_{-r}) \quad \dfrac{}{p_{1e} \vdash p_{1e}}\,(\mathrm{id})}{f_1, f_2, f_1 \otimes f_2 \multimap p_{1e} \vdash p_{1e}}\,(\multimap_{-l}) \quad \dfrac{}{f_3 \vdash f_3}\,(\mathrm{id})}{\cdots \quad f_1, f_2, f_3, f_1 \otimes f_2 \multimap p_{1e} \vdash p_{1e} \otimes f_3}\,(\otimes_{-r})}{\cdots \quad f_1, f_2, f_3, f_4, f_1 \otimes f_2 \multimap p_{1e} \vdash f_4 \otimes (p_{1e} \otimes f_3)}\,(\otimes_{-r})}{f_1, f_2, f_3, f_4, f_5, f_1 \otimes f_2 \multimap p_{1e} \vdash f_5 \otimes (f_4 \otimes (p_{1e} \otimes f_3))}\,(\otimes_{-r})$$

Figure 12
Proof of single dining philosopher sequent

We again omit the left branches of this proof consisting of identities $f_5 \vdash f_5$ and $f_4 \vdash f_4$.

After finishing his work (eating), p_{1e} releases both forks, i.e. t_{1e} is fired and we get the initial marking of PN, which is described by the sequent:

$$f_3, f_4, f_5, p_{1e}, p_{1e} \multimap (f_2 \otimes f_1) \vdash f_5 \otimes (f_4 \otimes (f_3 \otimes (f_2 \otimes f_1))), \tag{17}$$

and proven by the following (fragment of) proof in Figure 13:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{f_2 \vdash f_2}\,(\mathrm{id}) \quad \dfrac{}{f_1 \vdash f_1}\,(\mathrm{id})}{f_2, f_1 \vdash f_2 \otimes f_1}\,(\otimes_{-r})}{f_2 \otimes f_1 \vdash f_2 \otimes f_1}\,(\otimes_{-l}) \quad \dfrac{}{p_{1e} \vdash p_{1e}}\,(\mathrm{id})}{\cdots \quad p_{1e}, p_{1e} \multimap (f_2 \otimes f_1) \vdash f_2 \otimes f_1}\,(\multimap_{-l})}{\cdots \quad f_3, p_{1e}, p_{1e} \multimap (f_2 \otimes f_1) \vdash f_3 \otimes (f_2 \otimes f_1)}\,(\otimes_{-r})}{\cdots \quad f_3, f_4, p_{1e}, p_{1e} \multimap (f_2 \otimes f_1) \vdash f_4 \otimes (f_3 \otimes (f_2 \otimes f_1))}\,(\otimes_{-r})}{f_3, f_4, f_5, p_{1e}, p_{1e} \multimap (f_2 \otimes f_1) \vdash f_5 \otimes (f_4 \otimes (f_3 \otimes (f_2 \otimes f_1)))}\,(\otimes_{-r})$$

Figure 13
Proof of single dining philosopher after finishing work

Again, the missing left branches of the proof denoted by … contain identities for f_5, f_4 and f_3.

The second step is to enable two philosophers, e.g. p_2 and p_4, to eat as it is in Figure 15. In this step two transitions t_{2f} and t_{4f} can be fired simultaneously, because p_{2e} and p_{4e} have available both forks f_2, f_3 and f_4, f_5, respectively.

The corresponding linear logic formula:

$$f_1, f_2, f_3, f_4, f_5, (f_3 \otimes f_2 \multimap p_{2e}) \otimes (f_5 \otimes f_4 \multimap p_{4e}) \vdash f_1 \otimes (p_{4e} \otimes p_{2e}), \tag{18}$$

and its proof can be constructed in similar way as the proofs above depicted in Figure 14.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{f_2 \vdash f_2}\,(\mathrm{id}) \quad \dfrac{}{f_3 \vdash f_3}\,(\mathrm{id})}{f_2, f_3 \vdash f_2 \otimes f_3}\,(\multimap_{-l}) \quad \dfrac{\dfrac{}{p_{4e} \vdash p_{4e}}\,(\mathrm{id}) \quad \dfrac{}{p_{2e} \vdash p_{2e}}\,(\mathrm{id})}{p_{4e}, p_{2e} \vdash p_{4e} \otimes p_{2e}}\,(\otimes_{-r})}{\cdots \quad f_2, f_3, p_{4e}, f_2 \otimes f_3 \multimap p_{2e}, \vdash p_{4e} \otimes p_{2e}}\,(\multimap_{-l})}{f_2, f_3, f_4, f_5, f_4 \otimes f_5 \multimap p_{4e}, f_2 \otimes f_3 \multimap p_{2e} \vdash p_{4e} \otimes p_{2e}}\,(\multimap_{-l})}{\cdots \quad f_2, f_3, f_4, f_5, (f_4 \otimes f_5 \multimap p_{4e}) \otimes (f_2 \otimes f_3 \multimap p_{2e}) \vdash p_{4e} \otimes p_{2e}}\,(\otimes_{-l})}{f_1, f_2, f_3, f_4, f_5, (f_4 \otimes f_5 \multimap p_{4e}) \otimes (f_2 \otimes f_3 \multimap p_{2e}) \vdash f_1 \otimes (p_{4e} \otimes p_{2e})}\,(\otimes_{-r})$$

Figure 14
Proof of two dining philosophers

After finishing their dinner, the philosophers p_{2e} and p_{4e} release their forks, i.e. the transition t_{2e} and t_{4e} are fired simultaneously. The corresponding linear logic sequent and its proof can be constructed as above.

As the last step we consider that remaining philosophers p_3 and p_5 will eat. The corresponding linear sequent and its proof can be constructed similarly as in the previous steps. Now we have the solution of PN, because all philosophers had their dinner and we translated the behavior into corresponding provable linear sequents.

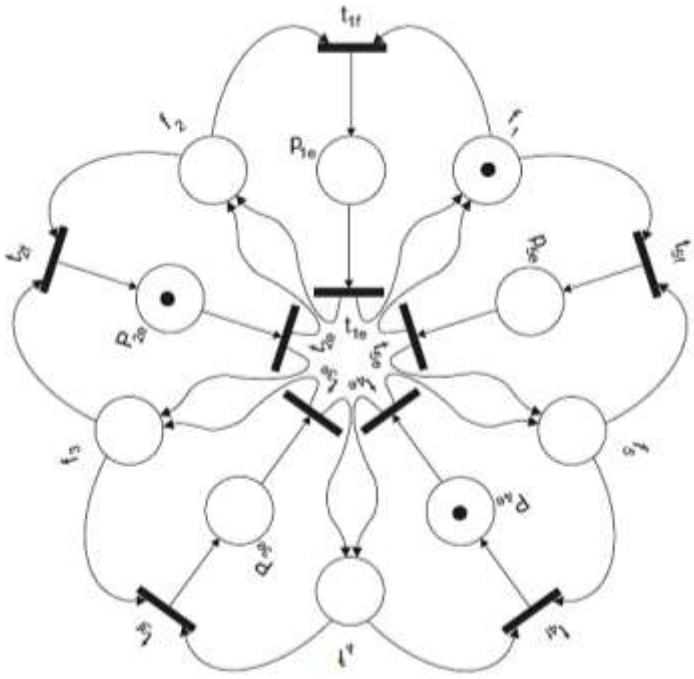


Figure 15
Two dining philosophers

Conclusions

In this paper, we presented an illustration on how the actual problem of component composition, used in component-based systems, can cause the advent of new problems. These problems need to be identified and solved. For achieving this aim, we apply a method of Petri nets transformation to the corresponding provable linear sequents. The proofs of constructed sequents ensure the correctness, they can be used as a specification of a system and they can help in achieving reliable software products. In our examples, we are concerned with only one kind of such problems, namely synchronization problems of mutual exclusion and deadlock that appear after the compositions of trivial working Petri nets. Our approach enables to formulate the solutions by corresponding provable sequents. We hope that our approach can be useful for education purposes, because it is simple, uses known concepts of PNs and linear logic, and is illustrative to comprehend component composition.

There are several open problems and possible ideas for solving them, by extending our approach. One of the advantages of linear logic is its resource character. First, we would like to formulate a transformation of timed Petri nets to linear formulas using polarization and focalization of the proof steps enabling to express time incrementally. After further study of Girard's theory of Ludics and its handling another resource, a space, we would like to transform explicit information about data types in colored Petri nets to linear logic proofs.

It is a challenge for us to work out our approach also for colored Petri nets, i.e. to define transformation from Petri nets, to the sequents of linear logic and illustrate it on significant examples.

The main result of our work is a systematically, worked out solution, for practical programmers, respectively, for students, on how different various known formal tools can be combined and proved, together with an illustration of component composition.

References

[1] Abramsky S.: Computational interpretations of linear logic, Technical Report 90/20, Department of Computing, Imperial College, London, pp. 1-15, 1990

[2] Aalst W. M. P., van der, Hee, K. M., van: Workflow management: models, methods, and systems, MIT Press, Cambridge, MA, 2002

[3] Aalst W. M. P., van der: The application of Petri nets to workflow management, J. of Circuits, Systems and Computations 8 (1), pp. 21-66, 1998

[4] Demeterová, E., Mihályi, D., Novitzká, V.: Component composition using linear logic and Petri nets, proceedings of the IEEE 2015 International Scientific Conference on Informatics, Poprad, pp. 91-96, 2015

[5] Engberg U., Glynn W.: Petri nets as models of linear logic, LNCS 431, pp. 147-161, 1990

[6] Girard J.-Y.: Linear logic, Theoretical Computer Science, Vol. 50, pp. 1-102, 1987

[7] Girard J.-Y.: Linear logic: Its syntax and semantics, Cambridge University Press, 2003

[8] Girard J.-Y., Taylor P., Lafont Y.: Proofs and types, Cambridge University Press, New York, NY, USA, 1989

[9] Girault C., Valk R.: Petri nets for systems engineering - a guide to modeling, verification, and applications, Springer, 2003

[10] Korečko, Š., Marcinčin, J., Slodičák, V.: A tool for management of networked simulations, Lecture Notes in Computer Science: Application and Theory of Petri Nets. Vol. 7347, pp. 408-417, 2012

[11] Korečko, Š., Sobota, B.: Petri nets to B-language transformation in software development, Acta Polytechnica Hungarica Vol. 11, No. 6, pp. 187-206, 2014

[12] Lincoln P.: Linear logic, SRI and Stanford University, 1992

[13] Martí-Oliet, N., Meseguer, J.: From Petri Nets to Linear Logics through Categories: A survey, World Scientific Publishing Company, International Journal of Foundations of Computer Science, Vol. 2, No. 4, pp. 297-399, 1991

[14] Novitzká V., Mihályi D.: Resource-oriented programming based on linear logic, Acta Polytechnica Hungarica, Vol. 4, No. 2, 2007, pp. 157-166

[15] Mihályi D., Novitzká V.: What about linear logic in computer science?, Department of Computers and Informatics, Technical University of Košice, Acta Polytechnica Hungarica, Vol.~10, No. 4, 2013, pp. 147-160

[16] Novitzká V. Mihályi D., Slodičák V.: Linear logical reasoning on programming, Acta Electrotechnica et Informatica, No. 3, Vol. 6, pp. 34-39, 2006

[17] Mihályi D., Novitzká V., Towards the knowledge in coalgebraic model of IDS, Computing and Informatics, Vol. 33, No. 1, 2014, pp. 61-78

[18] Perháč J., Mihályi D., Novitzká V.: Between syntax and semantics of resource oriented logic for IDS behavior description, Journal of Applied Mathematics and Computational Mechanics. Vol. 15, No. 2, pp. 105-118, 2016

[19] Perháč J., Mihályi D.: Intrusion Detection System Behavior as Resource-Oriented Formula, Acta Electrotechnica et Informatica. Roč. 15, No. 3, pp. 9-13, 2015

[20] Peterson, James L.: Petri theory and the modeling of systems, Englewood Cliffs, N.J.: Prentice-Hall, 1981

[21] Pietriková, E., Chodarev, S. Profile-driven source code exploration Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015: Lodz, Poland, pp. 929-934, 2015

[22] Polberger, D.: Component technology in an embedded system, Lund University, USA, 2009

[23] Raclet, J. B.: A modal interface theory for component-based design, Fundam. Informatics, Vol. 108, No. 1-2, 2011, pp.119-149

[24] Reisig, W.: Petri nets and algebraic specifications, Theoretical Computer Science 80 (1), pp. 1-34, 1991

[25] Riemann, R. C.: Structural and Semantical Methods in the High Level Petri Net Calculus, Herbert Utz Verlag, 1999

[26] Slodičák, V., Macko, P.: The role of linear logic in coalgebraical approach of computing, Journal of Information and Organizational Sciences. Vol. 35, No. 2, pp. 197-213, 2011

[27] Szyperski, C., Gruntz, D., Murer, S.: Component software beyond object-oriented programming, ACM Press, New York, 2002

[28] Vokorokos, L., Adám, N.: Secure web server system resources utilization, Acta Polytechnica Hungarica Vol. 12, No. 2, pp. 5-19, 2015

[29] Wang, A., Qian, K.: Component-oriented programming, Wiley-Interscience, 2005

[30] Winskel, G., Nielsen, M.: Models of Concurrency, In: Handbook of Logic and the Foundation of Computer Science, Vol. 4, 1993, pp.1-148

Assessment of the Sensory and Moisturizing Properties of Emulsions with Hemp Oil

Małgorzata Kowalska, Magdalena Wozniak, Magdalena Pazdzior

Kazimierz Pulaski University of Technology and Humanities, Chrobrego Street 27, 26-600 Radom, Faculty of Material Science, Technology and Design, Poland, e-mail: mkowalska7@vp.pl, m.wozniak@uthrad.pl, m.pazdzior@uthrad.pl

Abstract: The aims of this work were a sensory analysis, an evaluation of moistening properties and establishment of emulsions stability. Emulsions were prepared with two variable parameters (hemp oil content and time of homogenization). 15 respondents assessed emulsions sensory and skin moisturizing properties. Kleeman's optimizing method was used to designate the most stable emulsion system with the variable parameters. Taking into consideration all properties of the presented emulsions, emulsion V (50g of hemp oil and homogenized for 6 minutes) was found to be the optimum composition with the best moistening and sensory properties. After the application of this emulsion, the skin was highly moistened. Consistency of the emulsions was homogeneous and free from clotting. According to Kleeman's method, emulsion V was also confirmed as the most optimum variant of the emulsion. The work confirmed that hemp oil can be successfully used as a component for cosmetic emulsions.

Keywords: hemp oil; emulsions; sensory evaluation; moisturizing parameters; emulsions stability

1 Introduction

Emulsions constitute an important part in many fields of different industries, including the pharmaceutics, cosmetics, food and petrochemicals [1, 3, 12, 13]. In food industry emulsions can be used in low-calorie products, as a taste masking ingredients or improved sensory characteristics. In pharmaceutical industry – as a drug delivery systems, while in cosmetics – as a creams with encapsulated ingredients. [11, 19].

The prosperity of a cosmetic emulsion formulation depends among others on the stability, an efficiency of the active ingredients but highly important is also the consumer assent, which is definitely caused by the sensory characteristics of the product. Actually, the aims of the studies in cosmetic field are to make products

more innovative and successful, which will perfectly fulfill the sensory criteria of the customer. In general, customers choose a cosmetic emulsion for properties of active ingredients or effectiveness, but are mainly convinced by the positive feelings it brings to them, in particular in terms of texture. Therefore, sensory analysis dominates in the quality control of emulsions. This tool has rapidly been expanded and widely used to characterize and quantify texture properties of cosmetic dispersions [6, 14].

According to [15] sensory analysis is the evaluation of a product through the assessment of the properties perceptible by the following senses as taste, color, touch, texture, odor and noise, which results in receiving of the organoleptic profile of various products, for example cosmetics. Sensory evaluation interprets, measures and assess properties of a formulation, after stimulating users in connection to their main senses. So far, there is no instrumental methods, capable to estimate the human impressions, sensory analysis is a very effective method to receive data about a product in relation to consumer acceptance [4].

Vegetable oils, which contain unsaturated fatty acids, are frequently use in cosmetic dispersions to encourage customers and raise the sensory feelings. Hemp oil can be one of such oils and can also improve innovativeness of the cosmetic products. The attention directed to hemp for skin care is affected by great amount of oil in seeds of this plant (25-35%), but also a beneficial profile of unsaturated fatty acids consist of 50-60% linoleic acid (C18:2ω6), 20-25% α-linolenic acid (C18:3ω3) and a substantial portion of γ-linolenic acid (C18:3ω6). These three fatty acids constitute a structure of the cell membranes phospholipids. Moreover they have an effect on the immunology of the cells as well as on the several cell membrane functions such as the hormones activity, fluidity, or the transport of electrolytes [18, 20]. Linoleic and linolenic acids improve the skin structure and have a beneficial influence on dry and rough skin [20].

The aim of this work was to evaluate a sensory analysis and skin moisturizing properties of emulsions with hemp oil. To detect the most stable variant of emulsion, the optimization was performed using the software based on Kleeman's method [8]. The software indicated the variant with appropriate stability after only one measurement. It allows for reducing the time of the experiments and helps correctly design a stable dispersion.

2 Experimental

2.1 Materials

Emulsions were manufactured using the following components: distilled water, carboxymethylcellulose (Barentz Hoofddorp, Netherlands), cold pressed, unrefined (Gracefruit Ltd Stirlingshire, United Kingdom) (INCI: Cannabis Sativa Seed Oil) hemp oil containing the following fatty acids (FA): palmitic acid C16:0 – 5.7%, palmitoleic acid C16:1 – 0.15%, stearic acid C18:0 – 2.5%, oleic acid C18:1 9c – 11.4%, linoleic acid C18:2 9c12c – 55.4%, alpha linolenic acid C18:3 9c12c15c – 17.2%, gamma linolenic acid C18:3 6c9c12c – 3.6%, eicosenoic acid C20:0 – 0.6%, eicosenoic acid C20:1 – 0.4%, docosanoic acid C22:0 – 0.3%, tetracosanoic acid C24:0 – 0.2%, (Oleofarm company), sunflower lecithin (Lasenor, Emul, S.L. Barcelona, Spain) sodium benzoate (Orff Food Eastern Europe, Marki/Warsaw, Poland), aloe vera (FLP), Scottsdale, Arizona, United States, citric acid (Jungbunzlauer Basel, Switzerland).

2.2 Preparation of Emulsions

Lecithin (5.2%) was introduced to the oil phase (hemp oil). Carboxymethylcellulose (0.6%) was dispersed in distilled water, and then, aloe vera (0.2%) was added. Oil and aqueous phases were heated to 50–55°C in a water bath separately. Homogenization of the both phases was achieved by means of a high shear mixer at an equal speed of 36288 RCF for a suitable time (given in Table 1). Afterwards, the emulsions were cooled to room temperature, and sodium benzoate (0.25%) was added. Next pH was adjusted to 5.5 with citric acid using a pH-meter (Mettler Toledo) equipped with a calomel combined pH electrode. The content of changing components was presented in Table 1.

Table 1

Changing components/parameter of emulsions given according Kleeman's method

	Component [%w/w]/ Parameter	**EMULSION**					
		I	**II**	**III**	**IV**	**V**	**VI**
Changing components /parameter	Water	83.75	43.75	63.75	63.75	43.75	83.75
	Hemp Oil	10	50	30	30	50	10
	Mixing time [min]	1.5	1.5	3.0	4.5	6.0	6.0

2.3 Methods

2.3.1 Determination of Skin Capacitance

The determinations of skin capacitance were performed using a CM825 Corneometer (Courage+Khazaka Electronic). The principle of this method concerns the difference between capacitance of the dielectric constant of water and other substances. Each change of the dielectric constant subsequent to the modification of skin hydration results in an impaired calculated capacitance of a capacitor. The skin hydration degree values are in the range 0-130 arbitrary units (AU) [5].

In order to eliminate the influence of external conditions on the results, the measurements were performed under standard conditions (T° = 20-22°C, humidity 40-60%), away from direct sunlight. Women students of cosmetology or chemical technology (specialty: biotechnology and technology of cosmetics and household products) from the University of Technology and Humanities in Radom, Poland participated in those measurements.

The test included measurements carried out immediately prior to the application and after 15 min, 30 min, 45 min, 60 min, 90 min and 120 min after application of emulsion (approximately 0.01 g) on the designated forearm skin fragments. To have valid results, each measurement was taken three times. The region of the skin was wiped with a clean cotton cosmetic swab before and after the measurement. Variation in the time of the skin capacitance, as a difference between the values before and after application (as a mean value of 3 measurements for 15 respondents), were calculated by the following formula:

$$C = C_t - C_0 \tag{1}$$

where C_0 is a skin capacitance prior to the application and C_t is a skin capacitance over time t.

2.3.2 Sensory Determination

Sensory evaluation is based on the measurement and assessment of the product properties and consumer feelings by the senses (smell, taste, touch, sight).

Testing of the emulsion was performed by a group of respondents (15 females). The respondents were trained and specifically instructed on the methodology. A 5-point scoring scale was introduced, with 5 the maximum and 1 the minimum score. Details of the sensory assessments are presented in Table 2. The analysis was carried out at room temperature of 20°C (±2°C) and constant air humidity of 45% (±5%). Correctness of the testing was supervised by a researcher from the University of Technology and Humanities in Radom.

Table 2

Guidelines for sensory analysis of the cosmetic emulsions tested (study based on literature (Płocica et al. 2012; Płocica and Tal-Figiel 2009) and own experience (Kowalska et al. 2015)

Feature	Description of test procedure	Score (1–5)
C	Position hand at an angle of 60° and place 5 cm3 of the test substance there. Proceed to analyze its consistency by assessing the ability to keep the cosmetic adhering to the hand.	5. Cosmetic is easy to apply, not flowing 4. Easy to apply yet flowing can be observed 3. Cosmetic is hard to apply 2. Too thick to apply to the hand 1. Impossible to apply
H	Spread the substance on your hand and assess smoothness of its layer, presence of clots or air bubbles.	5. Completely homogeneous, no clots or air bubbles, forms a smooth layer on the skin 4. Homogeneous, no clots and few air bubbles, forms an uneven layer 3. Observable and palpable clots and air bubbles in the substance and on the skin when applied 2. Heterogeneous 1. Formulation components are not dissolved.
CE	Scoop 0.5 cm3 of the emulsion and rub between the thumb and index finger.	5. Imperceptible substance 4. Weakly perceptible substance 3. Somewhat perceptible substance 2. More perceptible substance 1. Highly perceptible substance
D	Spread 0.5 cm3 of the preparation on the forearm skin and observe its resistance to spreading.	5. No resistance to spreading 4. Little resistance to spreading 3. Incomplete cover, good spreading 2. Difficult to spread 1. Impossible to spread
SM	Apply 0.5 cm3 of the emulsion on the cleaned forearm skin and after an hour appraise the skin's smoothness in reference to a standard to which the substance has not been applied.	5. Very smooth, soft skin surface 4. Smoother and softer skin surface than of the reference standard 3. The skin surface is as smooth as that of the reference standard 2. Rough skin 1. Very rough skin
ST	Apply and spread the emulsion on the cleaned forearm skin, then press the other hand against this skin section and assess viscosity.	5. No palpable skin viscosity 4. Low skin viscosity 3. Palpable skin viscosity 2. Increased skin viscosity 1. High skin viscosity

G	Apply 0.5 cm3 of the substance on the cleaned forearm skin and assess formation of a greasy film	5. No sense of grease or film formation on the skin after application 4. Weak sense of greasiness, no film on the skin 3. Thin, greasy film on the skin after application 2. Greasy film on the skin directly on application 1. A compact, greasy film after application
A	Apply the substance on cleaned skin and assess the time of its absorption.	5. Very good absorption below 30 s 4. Good absorption from 30 s to 1 min 3. Average absorption from 1 to 3 min 2. Poor absorption from 3 to 5 min 1. Very poor absorption for more than 5 min

Legend:
- C - consistency (density and cohesion of the tested cosmetic),
- H - homogeneity (behavior of the preparation when applied to the skin – absence of clots or air bubbles)
- CE - cushion effect (palpability of the substance when rubbed between fingers),
- D - distribution (facility of spreading on the skin surface),
- SM - smoothing (smoothing effect when applied to the skin),
- ST - viscosity (degree of palpable viscosity left on the skin),
- G - greasiness (a fat film remaining on the skin),
- A - absorption (rate of absorption by the skin).

2.3.3 Determination of Viscosity

The viscosity of the emulsions were measured 24 h after manufacturing, at the speed of 10 rpm with spindle no. RV3, using a Brookfield Rheometer DV-I+. The measurements were performed at 25°C.

2.3.4 Determination of Mean Droplet Size and Droplet Size Distribution

The average droplet size and distribution were determined after 24 h using a Microtrac Particle Size Analyzer (Leeds & Northrup, Philadelphia, USA). Each measurement was repeated three times and given as the average value.

2.3.5 Dispersity Index

Dispersity index was calculated on the basis of laser diffraction droplet size measurements according to the formula:

$$k = (A - B)/C \quad (2)$$

where A, B and C are the biggest sizes of oil droplets for 90%, 10% and 50% of all particles, respectively.

2.3.6 Optimization of Parameters of Emulsion Stability

To obtain a stable emulsion optimization of parameters was carried out, using the Kateskór software, which is based on the Kleeman's method. The software was created for the Department of Technology Footwear and Tanning at the Faculty of Materials Science, Technology and Design (University of Technology and Humanities in Radom). The analysis was based on the measurement of viscosity and mean droplet size (input parameters) and as well our previous studies [9, 10] (Table 3). According to above mentioned factors a emulsion model was adopted, described by the following parameters for optimization purposes: dispersity index – 1.2, viscosity – 2800 mPa s, number of fractions – 1, average particle size – 5 μm, maximum particle size – 13 μm, minimum particle size – 0.5 μm.

Table 3
Values of viscosity and average particle size of emulsions measured after 24h from manufacturing

Emulsion	I	II	III	IV	V	VI
Average particle size [μm]	3.5	10.8	3.8	3.6	5.8	3.7
Viscosity [mPa*s]	200.0	4100.0	900.0	980.0	4800.0	110.0

3 Results

The assessment of the skin hydration degree carried out using a corneometer showed that all respondents had very dry skin [7]. The measurement of skin capacitance before application of each emulsion was in the range 25.2 – 28.6 AU. Figure 1 presents the difference between the skin capacitance (before and after application) related to time from application of the emulsions (as a mean value of 3 measurements for 15 respondents).

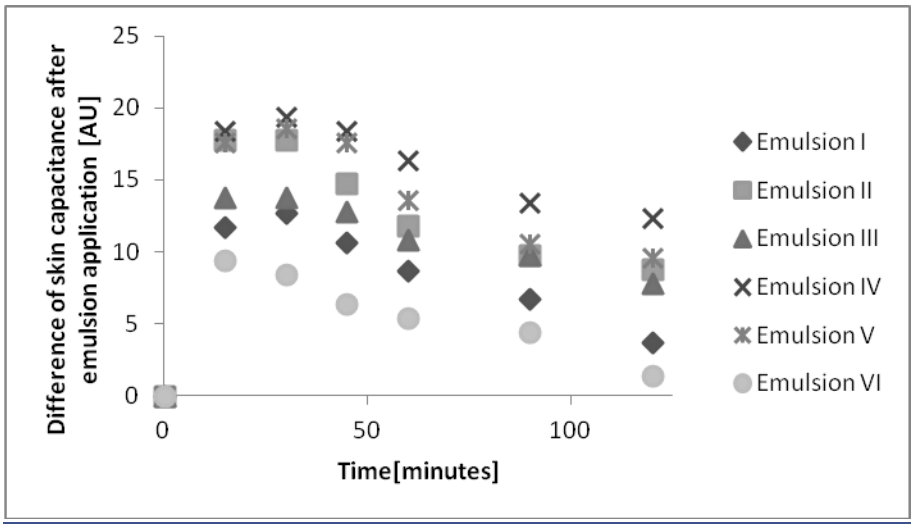


Figure 1
Variation in time of the skin capacitance after emulsions application as a difference between the values before and after application

After 15 minutes of application of each emulsion the corneometry values, compared to the previous measurements, significantly increased. The highest values were obtained after application of emulsions II, III, IV and V. While the lowest increase was recorded after application of emulsions I and VI. Thus, it can be assumed that the higher content of hemp oil in the emulsion caused higher degree of skin hydration. Obtained results clearly indicate a favorable effect of the high amount of hemp oil on the moisturizing properties of the examined emulsions. The corneometer showed a maximum increase of skin capacitance after 15 minutes to 50 minutes from the application of the emulsions. The highest increase of skin humidification was reported after application of emulsion V, which contained the highest amount of hemp oil and was homogenized the longest time. The proper homogenization (time and speed of homogenization) is the factor, which contributes to form an appropriate particle size of the emulsion and thus better penetration of the skin [2]. After 50 minutes from emulsions application the corneometry values decreased. In contrast, the Authors [5] obtained opposite results, because the maximum hydration was observed after one hour from application.

Sensory analysis provides the organoleptic evaluation of, among others, cosmetic products and is very helpful in providing the knowledge of how they are considered by the customers [15]. According to the results of the sensory analysis, emulsions II – V were evaluated the best, taking account the average value from all rated parameters. Generally, the respondents concluded, that emulsions had a

regular, homogeneous structure, good rate of absorption, smoothing effect proper value of greasiness.

Emulsions IV and V obtained the highest average score of all assessed parameters. However, the respondents the least highly valued the following parameters: cushion effect, smoothing, smoothing, greasiness and absorption (Table 4). Remaining parameters received 5 points (Figure 2). According to respondents, emulsions I and VI were the least satisfying (average scores 3.8 and 3.6 respectively). Cushion effect, consistency and distribution were rated the lowest (Table 4, Figure 2).

Generally, for the presented emulsions, sensory analysis confirmed significance of applying larger amount of oil to the emulsion system, as well as the importance of homogenization time for this type of formulations. Emulsions V and II contained an equal amount of oil, although homogenization time of emulsion II was much shorter (1.5 min), which affected to the fact, that this emulsion (II) did not fully meet with the respondents' expectations. Emulsion (IV), which contained smaller amount of oil (30 g), but homogenized for a longer time (4.5 min) was assessed better. It was observed, that particle size depends on homogenization time and oil content, which affects the stability and quality of emulsion systems [9].

Table 4

Mean values (± sd) of sensory evaluation received as a result of the survey

Parameter	Emulsion					
	I	**II**	**III**	**IV**	**V**	**VI**
CE*	3 ± 0.6	4 ± 0.7	4 ± 0.8	4 ± 0.6	5 ± 0.5	3 ± 0.9
H	5 ± 0.3	5 ± 0.2	5 ± 0.3	5 ± 0.2	5 ± 0.4	5 ± 0.6
C	3 ± 1.0	5 ± 0.5	4 ± 0.4	5 ± 0.4	5 ± 0.4	3 ± 0.9
D	3 ± 0.7	4 ± 0.7	4 ± 0.8	5 ± 0.3	5 ± 0.3	3 ± 0.4
ST	4 ± 0.3	5 ± 0.6	4 ± 0.2	4 ± 0.4	5 ± 0.2	4 ± 0.3
G	4 ± 0.9	4 ± 0.8	4 ± 0.4	5 ± 0.2	4 ± 0.7	4 ± 0.8
A	4 ± 0.8	4 ± 0.5	4 ± 0.7	4 ± 0.7	4 ± 0.7	3 ± 0.5
SM	4 ± 1.0	4 ± 0.5	4 ± 0.9	4 ± 0.9	5 ± 0.3	4 ± 0.7
Average	**3.8** ± 0.7	**4.4** ± 0.5	**4.1** ± 0.4	**4.5** ± 0.5	**4.8** ± 0.5	**3.6** ± 0.7

*See legend in Table II

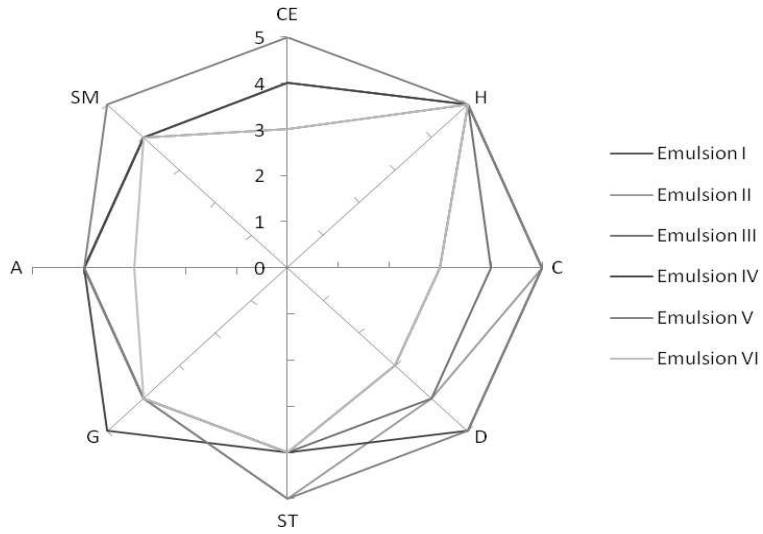


Figure 2
Sensory profile of emulsions (average values)

The optimization of the emulsions' stability demonstrated that the emulsions with two variables (the content of oil and time of homogenization) the optimum content of hemp oil should be in the range 30-50 g, while the optimum homogenization time should be in the range of 150 s to 360 s. Thus, emulsions indicated as optimal by the software were IV (contained 30 g of oil and homogenized for 4.5 minutes) and also V (contained 50 g of oil and homogenized for 6 minutes). Figure 3 presents a detailed study of the effect of the this two variables on the parameters of the emulsions.

Generally, it was observed, that increase of oil amount caused an increase of examined parameters' values. The exception was only dispersity index. In contrast, the increase of homogenization time leaded to increase of dispersity index and maximum particle size. However, the minimum and average particle size and viscosity remained practically unchanged (Fig. 3).

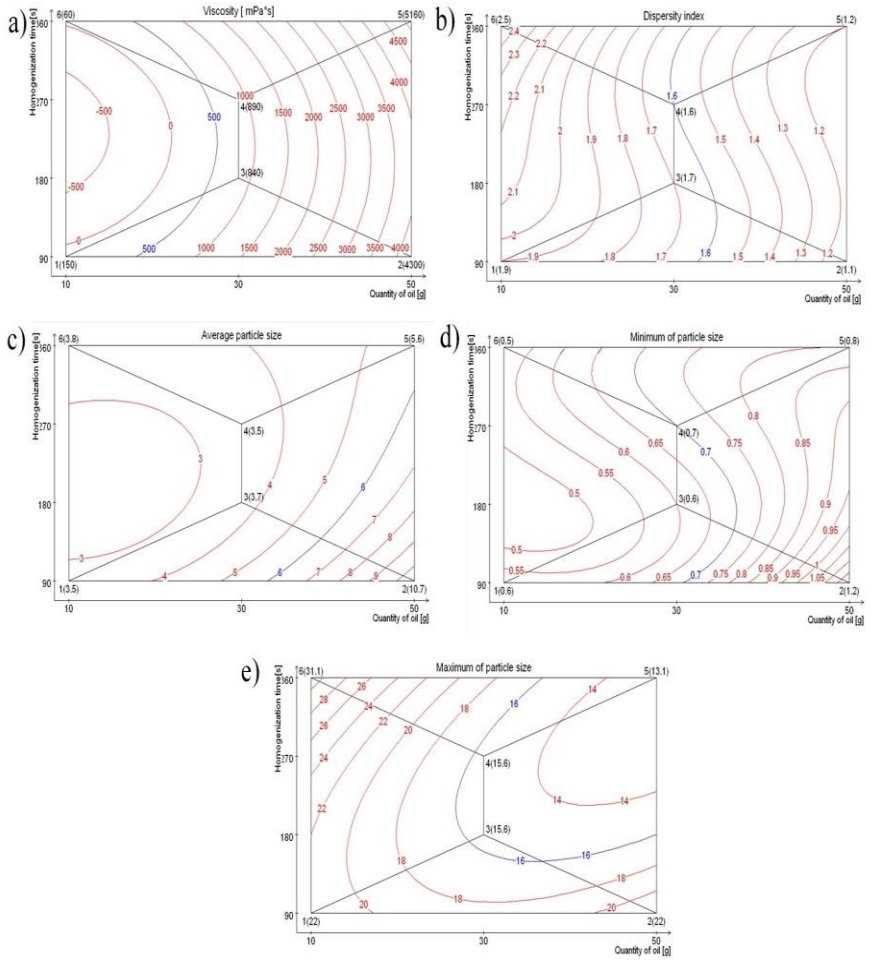


Figure 3

Effect of input parameters on: a) viscosity b) dispersity index c) average particle size d) minimum particle size e) maximum particle size

Conclusions

Hydration and improvements of the skin condition by the manufactured emulsions have been confirmed. Emulsions containing greater concentrations of hemp oil (50 g) and homogenized longer time (6.0 min) displayed maximum skin hydration. Consistency of the emulsion was homogeneous, smooth without clotting or air bubbles and was judged acceptable by the respondents. Additionally, all preparations retained a pleasant and had a delicate green color which makes such emulsions familiar for respondents.

The results corroborated the authors' hypothesis that hemp oil can be applied as an component of cosmetic emulsions with a low hydration.

The result obtained by computer analysis indicated that the most stable emulsion system is variant V (50 g of hemp oil, homogenized 6.0 min).

Acknowledgement

This work was supported by Kazimierz Pulaski University of Technology and Humanities.

References

[1] H. A. Abdulbari, N. H. Abdurahman, Y. M. Rosli, W. K. Mahmood, H. N Azhari, Demulsification of Petroleum Emulsions using Microwave Separation Method, International Journal of Physical Sciences 6(23), 5376-5382, 2011

[2] M. A. Bolzinger, S. Briançon, Y. Chevalier, Nanoparticles through the Skin: Managing Conflicting Results of Inorganic and Organic Particles in Cosmetics and Pharmaceutics. Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology 3(5), 463-478, 2011

[3] E. Bouyer, G. Mekhloufi, V. Rosilio, J. L. Grossiord, F. Agnely, Proteins, Polysaccharides, and their Complexes used as Stabilizers for Emulsions: Alternatives to Synthetic Surfactants in the Pharmaceutical Field? International Journal of Pharmaceutics 436(1), 359-378, 2012

[4] B. G. Chiari, M. A. Corrêa, M. G. J. De Almeida, V. L. B. Isaac, Cosmetics' Quality Control. INTECH Open Access Publisher (2012)

[5] M. M. Constantin, E. Poenaru, C. Poenaru, T. Constantin, Skin Hydration Assessment through Modern Non-Invasive Bioengineering Technologies Maedica 9(1), 33, 2014

[6] L. Gilbert, G. Savary, M. Grisel, C. Picard, Predicting Sensory Texture Properties of Cosmetic Emulsions by Physical Measurements, Chemometrics and Intelligent Laboratory Systems 124, 21-31, 2013

[7] U. Heinrich, U. Koop, M. C. Leneveu-Duchemin, K. Osterrieder, S. Bielfeldt, C. Chkarnat, H. Tronnier, Multicentre Comparison of Skin Hydration in Terms of Physical-, Physiological-and Product-Dependent Parameters by the Capacitive Method (Corneometer CM 825). International Journal of Cosmetic Science 25(1-2), 45-53, 2003

[8] W. Kleeman, Plast und Kautschuk. Das Leder. 11, 723, 1964

[9] M. Kowalska, A. Zbikowska, K. Śmiechowski, K. Marciniak-Lukasiak, Wpływ ilości lecytyny słonecznikowej i czasu homogenizacji na stabilność emulsji spożywczej zawierającej olej z orzechów włoskich. Żywność Nauka Technologia Jakość 21(1), 78-91, 2014

[10] M. Kowalska, M. Paździor, K. Śmiechowski, Wykorzystanie programu komputerowego opartego na metodzie Klemmana jako narzędzia pozwalającego uzyskać stabilną emulsję spożywczą. Postępy Techniki Przetwórstwa Spożywczego 24/44, 41-47, 2014

[11] M. Kowalska, M. Mendrycka, A. Zbikowska, S. Stawarz, Enzymatically Interesterified Fats based on Mutton Tallow and Walnut Oil Suitable for Cosmetic Emulsions. International Journal of Cosmetic Science 37(1), 82-91, 2015

[12] M. Kowalska, M. Ziomek A. Żbikowska, Stability of Cosmetic Emulsion Containing Different Amount of Hemp Oil. International Journal of Cosmetic Science *37*(4), 2015

[13] H. Mirhosseini, C. Tan, A. R. Taherian, Effect of Glycerol and Vegetable Oil on Physicochemical Properties of Arabic Gum-based Beverage Emulsion. European Food Research and Technology 228(1), 19-28, 2008

[14] T. Moravkova, P. Filip, The Influence of Thickeners on the Rheological and Sensory Properties of Cosmetic Lotions. Acta Polytechnica Hungarica 11, 173-186, 2014

[15] M. L. Piana, L. Persano Oddo, A. Bentabol, E. Bruneau, S. Bogdanov, C. Guyot Declerck, Sensory Analysis Applied to Honey: State of the Art. Apidologie *35*(Suppl. 1), 26-37, 2004

[16] J. Płocica, W. Figiel, B. Tal-Figiel, Sensory Research of Selected Products of Framona Company of the Line "Chocolate Delicacies". Polish Journal of Cosmetology 15, 134-137, 2012 (in Polish)

[17] J. Płocica, B. Tal-Figiel, Organoleptic Properties of Selected Cosmetics Containing Raw Plant Material from Sweet Clover and Rosemary. Polish Journal of Cosmetology 12, 67-76, 2009 (in Polish)

[18] G. Thews, E. Mutschler, P. Vaupel, Anatomie, Physiologie, Pathophysiologie des Menschen. Wissenschaftliche Verlagsgesellschaft, Stuttgart, Germany, 1999

[19] S. Van Der Graaf, C. G. P. H. Schroen, R. M. Boom, Preparation of Double Emulsions by Membrane Emulsification - a Review. Journal of Membrane Science 251, 7-15, 2005

[20] C. R. Vogl, H. Mölleken, G. Lissek-Wolf, A. Surböck, J. Kobert, Hemp (Cannabis sativa L.) as a Resource for Green Cosmetics: Yield of Seed and Fatty Acid Compositions of 20 Varieties under the Growing Conditions of Organic Farming in Austria. Journal of Industrial Hemp, 9(1), 51-68, 2004

An Experimental Study on the Phase Importance in Digital Processing of Speech Signal

Lazar Tesic, Boban Bondzulic, Milenko Andric, Boban Pavlovic

Military Academy, University of Defence in Belgrade, Generala Pavla Jurisica Sturma 33, 11000 Belgrade, Serbia
e-mails: lazar.tesic@vs.rs, boban.bondzulic@va.mod.gov.rs, milenko.andric@va.mod.gov.rs, boban.pavlovic@va.mod.gov.rs

Abstract: This paper presents the results of a late large-scale subjective study of phase importance of speech quality. Present study includes a collection of speech sentences distorted by limited transmission bandwidth and phase degradations. A detailed statistical analysis of the collected subjective judgments is presented. Among the used signal distortions, subjects preferred modification with signal phase preservation. Mean opinion scores of such modification are the closest to the original, undistorted sentences. This subjective study contributes to improving the algorithms of the speech processing, and in addition provide valuable data to develop objective or automatic methods of speech quality assessment, as well as to estimate their performance.

Keywords: phase spectrum; phase importance; speech signal; subjective evaluation

1 Introduction

Over the past decade listening devices and speech communication devices are more frequently used. Users of such devices expect their devices to provide good quality and intelligibility anywhere and at any time [5].

Most of the used digital processing approaches of speech signals exploit a short-time Fourier transform (FT). In this domain, signal is represented with complex-valued coefficients, which can, therefore, be observed by their magnitude and their phase. It is well known that for one dimensional and two dimensional signals, the magnitude and the phase of the FT play different roles in the signal reconstruction. Many authors emphasize that the phase of the FT is more important than the magnitude [2, 7, 11, 15-17]. The confirmation of this notion they mainly verified through signal modification in such way that phase is preserved, but the magnitude of all the spectral components is set to unity, i.e.

they did not consider end users (listeners or auditory) opinion. For example, authors in [2, 7, 15-17], illustrated the phase importance through phase-only image reconstruction. Moreover, authors of [11] employed objective speech quality measures for perceptual estimation of speech quality – signal-to-noise ratio (SNR) and perceptual evaluation of speech quality (PESQ) [8].

Justification of phase importance from a statistical viewpoint has been presented in [14]. It was shown that a random distortion of the phases can dramatically distort the reconstructed signal, while a random magnitudes distortion will not.

Most of the research papers, which study digital processing of speech signals through subjective quality assessment consider speech signal coding algorithms [9] and speech enhancement under noisy conditions [6, 19, 20]. International Telecommunication Union (ITU) coded-speech database [9] contains coded and source speech material used in the ITU-T 8 kbit/s codec. Authors in [20] have presented a Czech language speech database in a car environment (database contains signals in a quiet car without background noise, background noise in running car without speech and speech signals in running car). Paper [19] presented a database designed to evaluate the performance of speech recognition algorithms in noisy conditions (suburban train, crowd of people, car, exhibition hall, restaurant, street, airport and train station noises). Study reported in [6] shows performance of different speech enhancement algorithms on noisy speech corpus database (NOIZEUS). The same database has been used in [18], where the authors have shown that by modifying the phase spectrum in the enhancement process the quality of the resulting speech can be improved.

In the paper [22], the relation between uncertainty in phase and word error rate (WER) in human speech recognition through the subjective tests has been investigated. It has been shown that at an SNR of -10 dB, having random phases at all frequencies results in a WER of 63% compared to 24% if the phase was unaltered. With the SNR of 0 dB, random phase results in a 25% WER in respect to 11% for unaltered phase. It has been concluded that at high SNRs (i.e. 20 dB) the effect of phase on WER is small in comparison with low SNRs (such as 0, -5 and -10 dB), where the effect of phase on the recognition error rate can be significant.

Results from human listening tests [1] indicate that even for small window durations (20-40 ms), the phase spectrum can contribute to speech intelligibility as much as the magnitude spectrum.

Phase spectrum is used for quantifying speech quality in [4, 10]. In [10] authors proposed phase distortion deviation measure (PDD), which was evaluated in a database of dysphonic speakers with spasmodic dysphonia. They have shown that PDD is highly correlated with subjective ranking from medical doctors. Study [4] demonstrated that phase deviation objective metric is a reliable speech quality estimator, with performance in line with PESQ metric [8].

Good literature surveys about magnitude and phase importance in signal processing applications can be found in [13, 23]. While paper [23] presents a review on techniques for signal reconstruction without phase, paper [13] demonstrates the importance of phase in different applications including: speech enhancement, automatic speech, speaker recognition and speech synthesis.

This paper is focused on studying the phase importance in digital processing of speech signals through a large-scale subjective study. A human study was conducted using limited transmission bandwidth and phase degradations as speech sentence's distortions. Subjective trial was run at the University of Defence in Belgrade, Serbia, during the latter half of July 2015 and involved 18 listeners, evaluating 144 speech sentences. The listener responses gathered in the response directory at the end of trial were analysed and their basic statistics evaluated for each degraded sentence. Subjective trial provided an opportunity of evaluating listener performance within the context of speech observation.

The quality of speech as perceived by listeners is becoming increasingly important, due to the large number of listening and speech communication devices that humans utilize as the end users of speech.

The rest of the paper is organized as follows: Section 2 illustrates the magnitude- and phase-only signal reconstruction; in Section 3 the description of the used filters (signal degradations) is done; Section 4 presents the characteristics of the equipment and software tools used for signal acquisition; Section 5 describes the original and distorted sentences used, while Section 6 describes the subjective results and statistical analysis of the obtained results; finally, the conclusion is given in Section 7.

2 Reconstruction of Speech Signal Using Magnitude or Phase Spectrum

Information preservation about source (speech) signal in magnitude and phase spectrum is shown through example in Figure 1. The figure illustrates the waveforms of source signal (Figure 1(a)), signal reconstructed from the magnitude spectrum of source signal (Figure 1(b)) and the waveform of a signal produced using phase spectrum of source signal (Figure 1(c)). Reconstruction is performed by calculating inverse discrete Fourier transform (IDFT) over magnitude and phase spectrum of the source signal. When reconstructing from magnitude spectrum, it is assumed that the phase spectrum equals zero, and when reconstructing from the phase spectrum, it is assumed that the magnitude spectrum equals one. From Figure 1 it is obvious that the signal reconstructed from magnitude spectrum significantly differs from the source signal. Waveform of the source signal is preserved in the phase spectrum reconstructed signal.

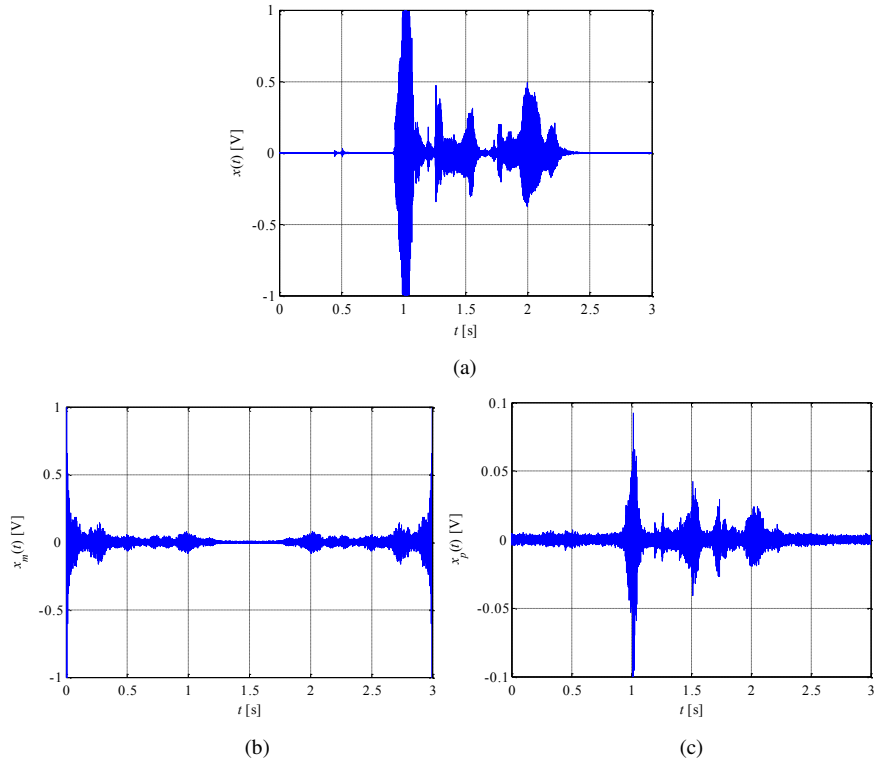


Figure 1
Waveforms of the: (a) source signal, (b) signal reconstructed from the magnitude spectrum of source signal and (c) signal reconstructed using the phase spectrum of source signal

3 Description of the Used Filters

For the purpose of phase spectrum importance research in digital signal processing of speech signals, verbal sentences are filtered with four types of low pass filters, which are marked as Type 1, Type 2, Type 3 and Type 4 (Figure 2). Type 1 represents the ideal low pass filter, with zero phase response. Type 4 is the real elliptic low pass filter. The remaining two types are produced by combining the magnitude ($\mu(f)$) and phase responses ($\theta(f)$) of the above-mentioned ideal and elliptic filters (Figure 2). Type 2 filter has ideal magnitude response and phase response taken from elliptic filter, and Type 3 filter has zero phase response and magnitude response taken from the elliptic filter.

Ideal low pass filter is designed for the cutoff frequencies of 1 kHz and 2 kHz, while the specifications of elliptic filter are given in Table 1.

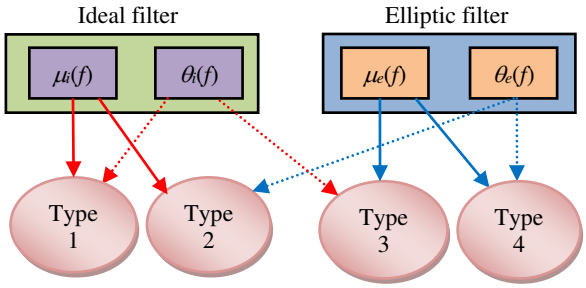


Figure 2
Filter design scheme

Table 1
Specifications for elliptic digital filter design

Cutoff frequency of the passband, f_g	1 kHz	2 kHz
Cutoff frequency of the stopband, f_a	2 kHz	3 kHz
Peak passband ripple	1 dB	
Minimum stopband attenuation	30 dB	
Sampling frequency	16 kHz	

The elliptic filter is an infinite impulse response (IIR) filter, and has a steep roll-off with equiripple in both passband and stopband. Using an elliptic filter design it is possible to achieve the lowest order for a given set of specifications [12]. The cutoff frequencies of projected filters are within a frequency range that contains the majority of the speech signal energy content (from 300 Hz to 2.8 kHz).

Elliptic function computations carried out by the following MATLAB® functions: (1) `ellipord` – for elliptic filter order selection and (2) `ellip` – which returns the filter coefficients.

The transfer functions $H(z)$ of the filters designed according to the specifications in Table 1 are given by:

$$H_{1\,\mathrm{kHz}}(z)=\frac{0.0274-0.0130z^{-1}-0.0130z^{-2}+0.0274z^{-3}}{1-2.5107z^{-1}+2.2227z^{-2}-0.6833z^{-3}} \tag{1}$$

$$H_{2\,\mathrm{kHz}}(z)=\frac{0.0529-0.0467z^{-1}+0.0940z^{-2}-0.0467z^{-3}+0.0529z^{-4}}{1-2.6643z^{-1}+3.2092z^{-2}-1.9105z^{-3}+0.4849z^{-4}} \tag{2}$$

Equations (1) and (2) show that the given specifications can be realized with elliptic filters of 3rd (f_g=1 kHz) and 4th order (f_g=2 kHz).

Locations of poles (x) and zeros (o) of the designed elliptic filters, Eqs. (1) and (2), are illustrated in Figure 3. The designed filters are stable because their poles are within the unit circle [3, 21, 24].

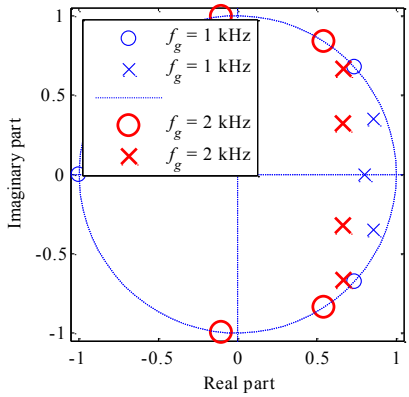
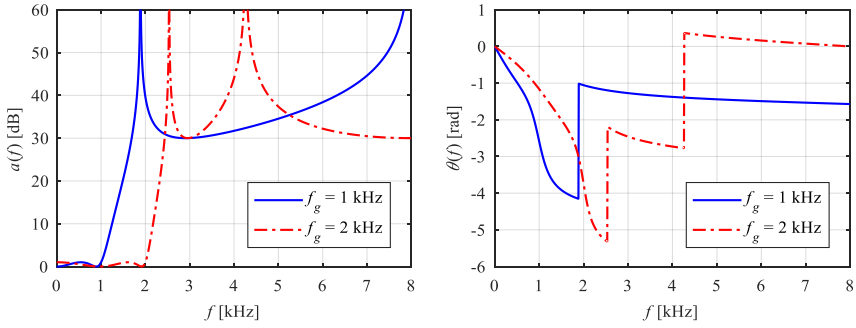


Figure 3

Locations of poles and zeros in the z-plane of the elliptic filters designed according to the specifications in Table 1

The magnitude and phase responses of the designed elliptic filters are given in Figure 4. Figure 4(a) shows that the attenuation requirements – peak passband ripple and minimum stopband attenuation, are satisfied. The phase responses of designed filters, Figure 4(b), are nonlinear. The nonlinear phase response resulting in group delays, *gd*, that are not constant in the passbands of the filters (Figure 5), which means that the components that enter the filter will not be delayed by the same value of time on its exit.



(a) (b)

Figure 4

(a) magnitude responses and (b) phase responses of the filters designed according to the specifications given in Table 1

An ideal filter (it is ideal in the sense that it is not realizable) is noncausal, and it was applied by calculating discrete FT (using MATLAB® function `fft`) of the input signal and by leaving all frequency components of a signal below a designated cutoff frequency (all components above are rejected).

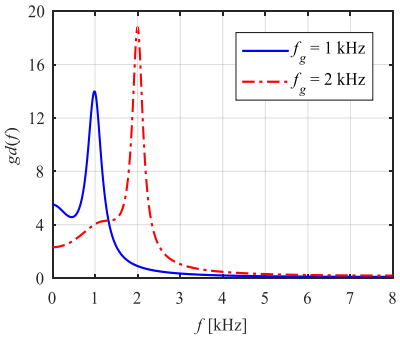


Figure 5

Group delays of the filters designed according to the specifications given in Table 1

4 Characteristics of the Equipment and Software Tools Used for Signal Acquisition

Test signals used in this research are audio – speech sentences. For speech acquisition Genius HS–400A Headband PC Headset with rotating microphone is used. Its features are shown in Table 2 and Table 3.

Table 2

Headphone features

Wearing style	on ear
Driver unit	40 mm
Frequency response	20 Hz – 20 kHz
Impedance	32 Ω
Sensitivity	102 dB
Cable length	1.8 m
Volume control	yes
PC in	2x3.5 mm stereo jack

Table 3

Microphone features

Sensitivity	-54+/-3 dB
Directivity	omni-direction
Frequency response	100 Hz – 10 kHz

PC used for recording and signal processing is a laptop computer Acer Aspire 5750G. Software used for signal processing was MATLAB® – version R2013a. A/D signal conversion was done with sampling frequency of 16 kHz, with 16-bit resolution.

5 Forming the Speech Database and Running the Subjective Tests

The phase importance research was realized through four phases: (1) gathering source speech sentences, (2) forming the speech database, (3) subjective tests running, (4) results arrangement and analysis.

First phase included source speech sentences recording. Three native speakers (two male and one female) were reading six Serbian sentences each (translation in English is in brackets):

"Aba je najbolji bend." (Abba is the best band.)

"Ada je recno ostrvo." (Ada is a river island.)

"Igi Pop je muzicar." (Iggy Pop is a musician.)

"Op'o mi pritisak." (My blood pressure went down.)

"Oko mi je crveno." (My eye is red.)

"OTO je predmet u osnovnoj skoli." (OTO is a subject in elementary school.)

Every sentence was recorded. Hardware and software tools used for signal recording and processing were described above.

In this way 18 source signals were recorded (sources-originals-references) lasting 3 seconds each. These sentences are the signals to be filtered in the next research step.

These sentences meet the ITU conditions [8]. Namely, sentences should be formed in that manner so that the six most frequent consonants (b, d, g, p, k, t), should be placed between two same vocals.

In the second research phase test speech sentences were formed (degraded sentences). Eighteen signals mentioned before were filtered with four types of filters described earlier. Filters were projected for two different cutoff frequencies, so every source sentence was modified in 8 different ways (4 types of filters x 2 cutoff frequencies). In this way, 144 test sentences were produced. So, the complete database contains 162 speech recordings (18 originals + 144 degraded signals), and they represent the ground basis for subjective testing. Using different types of filters and different cutoff frequencies, different signal quality is achieved. Therefore, different grades are expected.

Procedures and standards for subjective evaluation of speech signals have existed for many years. Because of the variety of digital audio contents in general, running subjective evaluation tests, for estimating audio quality, became very simple.

Quality estimation of test speech sentences was performed through subjective tests (third research phase). Subjective testing was performed on a representative

sample of several listeners. The testing followed ITU recommendation [8], which defined a phonetic approach in subjective testing of speech signals. Test signals must be produced by male as well as female speakers. Recommended duration of sentences is 1 to 3 seconds, where pure speech must be in a range of 40 to 80%. Minimum sampling frequency is 8 kHz and resolution should be 16 bits. Optimal number of listeners is 12.

Subjective tests included 18 listeners. The process went like this: the original sentence is played first, and the listener knew that, so he/she didn't evaluate the first sentence. After that, four respective degraded sentences are played (with cutoff frequency of 1 kHz of the passband) and at the end the original is played again, and the listener didn't know that he listens to the original (so he/she evaluated the original). All five sentences are graded with scores from 1 to 5 (1 – bad quality, 5 – excellent quality). During evaluation, speech comprehension sound purity, sound loudness and the sentence tone are taken into account. When the first round ends, sentences at the cutoff frequency of 2 kHz of the passband are listened to again and evaluated.

Each listener had to evaluate 27 different sentences (24 degraded and 3 originals). The listener, firstly, listened to the three originals from one speaker (and its respective degraded sentences), and then listened to subsequent three originals, and so on. Degraded sentences are randomly played, as suggested in [9].

6 Results and Analysis

The outcomes of subjective tests are quality-grades given by listeners, which are presented through mean opinion scores (MOS). MOS is the most commonly used method of generalising a subjective score given by a number of independent observers with respect to perceived quality of a signal. Extensively used in both subjective audio (in particular voice over IP (VoIP) communications) and video quality evaluation, MOS is simply the arithmetic mean of all individual scores assigned by the listeners (test participants) to a signal:

$$MOS_i = \frac{1}{N_S} \sum_{n=1}^{N_S} SQ(n,i) \tag{3}$$

where are:

i – degraded sentence index,

$SQ(n,i)$ – subjective grade given by n-th listener to i-th sentence and

N_s – the number of listeners who graded i-th sentence.

MOS is within the same range as the quality scale adopted during the subjective trials so there is no need for additional normalisation of the range. This is most commonly 0 or 1 to 5, as is the case with subjective trials ran here where 1 means the lowest and 5 the highest quality. MOS is a democratic measure in that it treats each subjective vote equally and the only true mean opinion.

Since subjective MOS scores are constructed from several individual quality scores, their estimate of absolute video quality has a statistical uncertainty associated with it. In cases where individual subject quality scores vary widely this uncertainty is large. Subjective score uncertainty can be measured using a number of methods, but the benchmark metrics of uncertainty are standard deviation, or its square the variance, and standard error. Standard deviation is evaluated directly from the individual subjective scores for sentence i, SQ_i (index n of the listener used previously is ignored here for the sake of brevity) and the MOS for that speech signal:

$$\sigma_i = \sqrt{\mathrm{E}\left[SQ_i^2\right] - \left(\mathrm{E}\left[SQ_i\right]\right)^2} = \sqrt{\mathrm{E}\left[SQ_i^2\right] - MOS_i^2} \tag{4}$$

where $E[X]$ is expected value or an average score of the random variable. Standard deviation is mostly shown together with an average score as its positive and negative variation.

The alignment of quality scores vary from listener to listener. Since listeners are free to choose which grade to assign, it is natural to expect that grades given to sentences differ from one listener to another. More generous listeners provide higher scores, while those who are stricter choose lower ones. Although the absolute range of the subjective scores doesn't influence the speech signal ranks, contradictory scores may influence MOS.

Equations (3) and (4) show the way of determining average grades and standard deviations of subjective scores of a single sentence (i-th sentence), which listeners evaluated. In this paper, the analysis of gathered scores was done through calculating average scores and standard deviations of subjective grades given to test sentences which are produced from the same source sentence, subjective scores of the sentences which were listened by every listener alone, subjective scores of test sentences which are produced by every speaker alone and subjective scores associated to different modifications (filter types) of a source speech sentences.

Figure 6 shows MOS scores histogram with 10 equally spaced bins – vertical axes corresponding to the number of sentences in each bin. We can see that the subjective trial contained a good spread of speech quality as mean opinion scores show variation of around 75% of the entire score range (from 2 to 5). Histogram maximum is in the medium quality area, wherein the dissipation of subjective scores around the maximum can be approximated with Gaussian (normal) distribution.

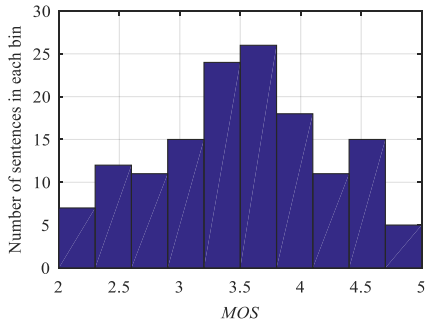


Figure 6
MOS scores histogram with 10 equally spaced bins

Figure 7 shows MOS values with respective reliability intervals (MOS±σ) of subjective scores of test sentences which are produced from the same source sentence (3 speakers x 6 sentences = 18 source sentences). A group of source sentences with the ordinal numbers 1 to 6 originating from the first speaker, group of sentences with the ordinal numbers from 7 to 12 originating from the second speaker, and the last group of six sentences originating from the third speaker. MOS values are within the range from 3.25 (sentence number 3) to 3.79 (sentence number 13), while standard deviations are within the range of 0.78 (sentence number 10) to 1.06 (sentence number 12).

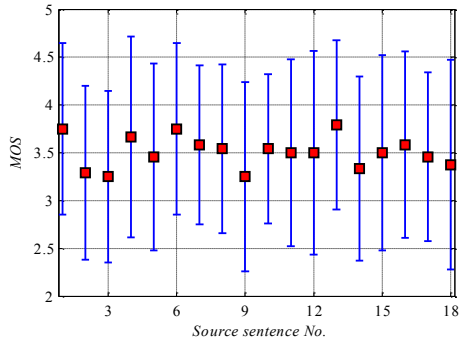


Figure 7
MOS values with respective reliability intervals of subjective scores of test sentences which are produced from the same source sentence

Listener performance is another important indicator of subjective trial success and quality/usefulness of resulting results. With respect to individual listeners' opinions we are primarily interested in analysing the consistency of their scoring. Obviously trials with high levels of agreement between individual subjects are more useful as the certainty of the resulting subjective quality scores is high. Conversely, in cases where subjects disagree and we have a large variance in

subjective quality and ranking we cannot be certain which speech sentence really has a quality advantage or it if the advantage really exists. In reality agreement between subjects is never ideal and a certain level of uncertainty remains for both absolute quality and ranking based on quality. The simplest approach to measure general subject performance within a trial is to evaluate some basic statistics of their quality responses.

The basic statistics of listeners' subjective scores (average scores and standard deviations) are shown on Figure 8. Average values are within a range from 3.3 to 3.83, with relatively equal reliability intervals. Although the difference between the MOS values exist, it doesn't point out one listener from the majority.

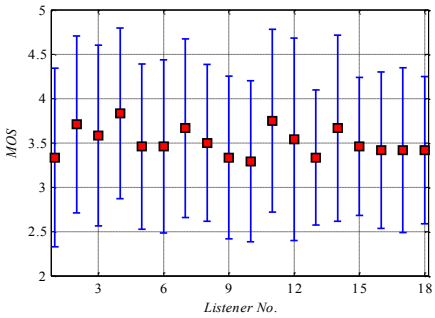


Figure 8

MOS values with respective reliability intervals of subjective scores given to sentences by one listener

Figure 9 shows MOS values with reliability intervals of subjective scores of test sentences which came from the same speaker (three speakers). It is obvious that the mean scores are relatively equal and that they are about 3.5. Standard deviations are also very close, and they are between 0.916 (speaker 2) and 0.968 (speaker 3).

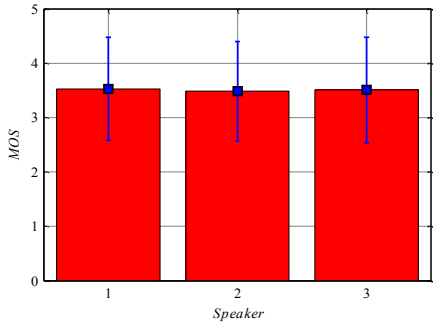


Figure 9

MOS values with respective reliability intervals of subjective scores of test sentences which came from the same speaker

To analyse the ranking of various degraded sentences the results are best viewed on a filter level. The scores are aggregated according to filter and shown on Figure 10, which shows MOS values of subjective scores given to the sentences with the same type of degradation – filter type analysis. Filter types are described in Section 3. Beside the average scores of the test sentences, Figure 10 shows the average value of scores given to the source sentences (without degradation).

Figure 10 shows that the biggest MOS value belongs to the originals – source sentences (4.85), which was expected. Additionally, listeners evaluated that better quality is maintained when sentences passed through filters Type 3 and 4 then through filters Type 1 and 2. This was also expected because filters Type 1 and 2 have ideal magnitude responses (both with cutoff frequencies of 1 kHz or 2 kHz), while filters Type 3 and 4 have magnitude responses with transition zones 1 kHz wide (both types include transition zones from 1 kHz to 2 kHz or from 2 kHz to 3 kHz). Transition zone enables for the spectral components which are higher than cutoff frequency of the passband, not to be completely attenuated, i.e. to be perceptually noticeable.

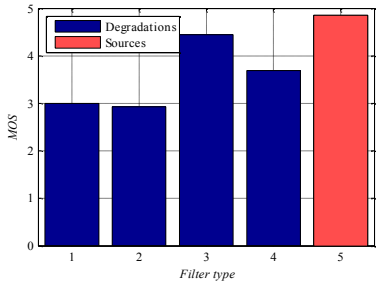


Figure 10

MOS values of subjective scores given to the sentences with the same type of degradation

Test sentences which are modified with Type 3 filter (magnitude response is the same as elliptic filter and phase response is zero) are graded with the highest average score of 4.44. Phase response of this filter is set to zero, so, input and output signals have the same phase spectrums, i.e. the original signal phase is completely preserved. The importance of phase preservation is obvious when MOS values of the filters of the same magnitude responses are compared (Type 1 versus Type 2, or Type 3 versus Type 4). In both cases, zero phase filters have the priority (Type 1 and Type 3). The benefit in phase preservation is much greater when analyzing the filter with non-ideal (real) magnitude response (MOS_3=4.44, MOS_4=3.69) and comparing it to the filter with ideal magnitude response (MOS_1=3, MOS_2=2.92). This result can be interpreted with the phase preservation influence – when analyzing filters with ideal magnitude response, phase response can be analyzed only in the passband, while when analyzing filters with real magnitude response the phase influences the transition zone and the stopband band as well (spectral components of the original signal are not completely attenuated).

Figure 11 shows MOS values of subjective scores given to the test sentences with the same type of degradation, when different cutoff frequencies of the passband of designed filters are taken into analysis separately.

Comments which refer to the source signal degradation analysis, when analyzing the complete database (Figure 10), are all in line when analyzing the source signal degradation with both cutoff frequencies separately (Figure 11). Additionally, MOS values of degraded sentences provided through filters with 2 kHz cutoff frequency of the passband (Figure 11(b)) are higher than MOS values of degraded sentences provided with filters with 1 kHz cutoff frequency of the passband (Figure 11(a)).

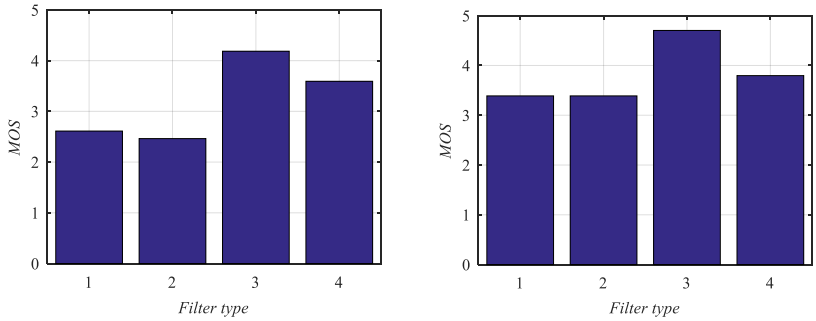


Figure 11

MOS values of subjective scores given to the test sentences with the same type of degradation: (a) for the cutoff frequency f_g=1 kHz and (b) for the cutoff frequency f_g=2 kHz

The analysis results indicate that the test material was correctly chosen and prepared; also the subjective tests were well done. Because of that, the test sentence database with the entire subjective test results can be used for development of objective quality estimation algorithms for speech (audio) signals. The idea is to compare directly, MOS quality values with the values gained from the objective quality estimation algorithms.

We have decided to make the test material available to the research community free of charge [25]. Along with speech sentences and subjective scores, we provided MATLAB® files, too.

Conclusions

A subjective study to evaluate the phase importance on the perceptual quality of speech communication was presented. This study included 144 speech sentences derived from 18 original sentences using eight distortion types and were evaluated by 18 listeners. Subjective quality data collected through subjective trials was processed into the form of mean opinion scores expressing mean estimates of speech quality for each degraded sentence used in the study. The resulting

database is unique in terms of content and distortion and is publicly available to the research community for further research on speech quality assessment.

For the purpose of validation, a specific set of software tools was constructed for conducting the validation and performing the comparison between subjective quality scores. Specifically, a set of statistical tools was designed and implemented in Matlab® development environment that allows reading, comparison and output of a set of both quantitative and qualitative validation scores.

The results of the performed study show that in the modeling and processing of the time-frequency signal representation, phase can't be ignored. Subjective results may with further analysis provide deeper insight into how people decide and what influences the decisions they make regarding perceived speech quality.

In future work the speaker database should be extended. Furthermore, we will develop objective speech quality assessment measure, with special attention to phase preservation measuring.

Acknowledgement

This research has been a part of the project No. VA-TT/1-17-19 supported by the Ministry of Defence, Republic of Serbia.

References

[1] Alsteris, L. D., Paliwal, K. K.: Further intelligibility results from human listening tests using the short-time phase spectrum, Speech Communication, 2006, Vol. 48, No. 6, pp. 727-736

[2] Behar, J., Porat, M., Zeevi, Y. Y.: Image reconstruction from localized phase, IEEE Transactions on Signal Processing, 1992, Vol. 40, No. 4, pp. 736-743

[3] Besancon-Voda, A.: Iterative auto-calibration of digital controllers: methodology and applications, Control Engineering Practice, 1998, Vol. 6, No. 3, pp. 345-358

[4] Gaich, A., Mowlaee, P.: On speech quality estimation on phase-aware single-channel speech enhancement, In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane (Australia), 2015, pp. 216-220

[5] Gerkmann, T., Krawczyk-Becker, M., Le Roux, J.: Phase processing for single-channel speech enhancement: History and recent advances, IEEE Signal Processing Magazine, 2015, Vol. 32, No. 2, pp. 55-66

[6] Hu, Y., Loizou, P. C.: Subjective comparison and evaluation of speech enhancement algorithms, Speech Communication, 2007, Vol. 49, No. 7, pp. 588-601

[7] Huang, T. S., Burnett, J. W., Deczky, A. G.: The importance of phase in image processing filters, IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, Vol. ASSP-23, No. 6, pp. 529-542

[8] ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ), An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, February 2001

[9] ITU-T Telecommunication standardization sector of ITU, Series P: Telephone transmission quality, telephone installations, local line networks, Supplement 23: ITU-T coded-speech database, February 1998

[10] Koutsogiannaki, M., Simantiraki, O., Degottex, G., Stylianou, Y.: The importance of phase on voice quality assessment, In Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH), Singapore, 2014, pp. 1653-1657

[11] Langarani, M. S. E., Veisi, H., Sameti, H.: The effect of phase information in speech enhancement and speech recognition, In Proceedings of the 11th International Conference on Information Sciences, Signal Processing and their Applications (ISSPA), Montreal (Canada), 2012, pp. 1446-1447

[12] Mitra, S. K.: Digital Signal Processing – A Computer Based Approach, 3rd ed., New York (USA): McGraw-Hill, 2006, ISBN: 007-124467-0

[13] Mowlaee, P., Saeidi, R., Stylianou, Y.: INTERSPEECH 2014 Special Session: Phase importance in speech processing applications, In Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH), Singapore, 2014, pp. 1623-1627

[14] Ni, X.S., Huo, X.: Statistical interpretation of the importance of phase information in signal and image reconstruction, Statistics & Probability Letters, 2007, Vol. 77, No. 4, pp. 447-454

[15] Oppenheim, A. V., Hayes, M. H., Lim, J. S.: Iterative procedures for signal reconstruction from Fourier transform phase, Optical Engineering, 1982, Vol. 21, No. 1, pp. 122-127

[16] Oppenheim, A. V., Lim, J. S.: The importance of phase in signals, Proceedings of the IEEE, 1981, Vol. 69, No. 5, pp. 529-541

[17] Oppenheim, A. V., Lim, J. S., Kopec, G., Pohlig, S. C.: Phase in speech and pictures, In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Washington, D.C. (USA), 1979, Vol. 4, pp. 632-637

[18] Paliwal, K., Wojcicki, K., Shannon, B.: The importance of phase in speech enhancement, Speech Communication, 2011, Vol. 53, No. 4, pp. 465-494

[19] Pearce, D., Hirsch, H.-G.: The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy

conditions, In Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP), Beijing (China), 2000, Vol. 4, pp. 29-32

[20] Pollak, P., Vopicka, J., Hanzl, V., Sovka, P.: CAR2 – Czech database of car speech, Radioengineering, 1999, Vol. 8, No. 4, pp. 1-6

[21] Precup, R.-E., Preitl, S.: PI and PID controllers tuning for integral-type servo systems to ensure robust stability and controller robustness, Electrical Engineering, 2006, Vol. 88, No. 2, pp. 149-156

[22] Shi, G., Shanechi, M. M., Aarabi, P.: On the importance of phase in human speech recognition, IEEE Transactions on Audio, Speech and Language Processing, 2006, Vol. 14, No. 5, pp. 1867-1874

[23] Sturmel, N., Daudet, L.: Signal reconstruction from STFT magnitude: A state of the art, In Proceedings of the 14th International Conference on Digital Audio Effects (DAFx), Paris (France), 2011, pp. 375-386

[24] Tar, J. K., Bitó, J. F., Rudas, I. J.: Contradiction Resolution in the Adaptive Control of Underactuated mechanical Systems Evading the Framework of Optimal Controllers, Acta Polytechnica Hungarica, 2016, Vol. 13, No. 1, pp. 97-121

[25] Tesic, L., Bondzulic, B., Andric, M., Pavlovic, B.: An experimental study on the phase importance in digital processing of speech signal – Phase importance (PhImp) database [Online]. Available: https://mega.nz/#!WsMEhTJQ!IHtEQ4XQ_Ir-tRPU-qqbmf0xcqNsl2TWmTJXt_PWCnQ, October 2017

Local Muscular Load Measurement with the Help of a Datalogger

David Tuček, Barbora Dombeková

Tomas Bata University in Zlin
nam. T.G. Masaryka 5555, 760 01 Zlin, Czech Republic
E-mail: tucek@fame.utb.cz, dombekova@fame.utb.cz

Abstract: The aim of this research was the development and verification of the local muscular load measuring methodology as a valid and recognized measurement in order to subsequently strengthen injury prevention of occupational diseases among workers of both manufacturing and non-manufacturing companies. A milestone in the project was construction of a measuring device to obtain data for evaluating local muscular load with the use of force measurement, frequency of movements and positions and a device called datalogger. This measuring device was then tested and its functionality has been successfully proved in 40 measurements in both manufacturing and non-manufacturing enterprises. The reason for the development of this equipment was especially a requirement to learn the load (force generation) of the individual fingers when operators work, implemented by different grips, e. g. tridigital, pentadigital and so on. Such quantification is not possible with the current EMG method.

The developed equipment had to be validated with existing certified methodology called electromyography, which allows only summative evaluation. The results of correlations of both measurements while evaluating the local muscular load are analyzed in this article. The aim of this paper is also to describe how the new device was developed and tested: the Datalogger for measuring local muscular load and its advantages and limitations. Based on the defense carried out, the Industrial Property Office subsequently issued a letters patent for the invention with a number PV 2015-820 earlier this year. The invention relates to an ergonomic device developed for monitoring of local muscular load, and is especially useful in workplaces where there is increased stress, particularly stress on your hands when multiple / repetitive movements, often involving spending increased muscle strength. Measurements can also be made for operators wearing different preservative gloves commonly used in practice.

In this article the authors will introduce not only new patented apparatus itself, but a new methodology for local muscular load measuring with the datalogger as well.

Keywords: ergonomics; local muscular load, carpal tunnel syndrome; datalogger; long-term, integrated electromyography (EMG)

1 Introduction

We are definitely at the beginning of Industry 4.0. Industry 4.0 is the next phase in digitalization of the manufacturing sector. Most of the digital technologies have been brewing for some time. Some are not yet ready for application at scale [1]. Industry 4.0. is based on informatics, technical and autonomous pillars. But, in our article we focused on human factor. The human factor it is the most important factor in companies. Therefore, it is essential to create healthy and supportive working conditions. Otherwise, the employee risks exposure to occupational diseases, serious illnesses, etc. In this case, the employer is then exposed to liabilities stemming from not creating suitable working conditions, such as compensation, increased recruitment costs, etc. In this context, the very important role is played by social environment and political and legal environment that is created by the state authorities [2].

In the Czech Republic and Slovak Republic management is being increasingly responsible for occupational diseases stemming from "local muscular load" of employees. In the rest of the world it is also known as the repetitive strain injury (RSI).

We will compare results of our measuring with those that we achieved by traditional certified apparatus for the Local Muscular Load – Electromyography (EMG holter). The reliability is possibly expressed by a relative reliability coefficient, and we study it by using several methods, e.g. parallel test measuring, which means a measurement conformity with other adequate measurements of the same construct. In our procedure, a maximum reliability is offered thanks to using the EMG holter, which is then studied by algorithm of searching for the highest conformity – correlation.

2 Ergonomics – Specifics in the Czech and Slovak Republics

Ergonomics is the discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance" [3, 4]. At the same time, suitable ergonomic solution is the way to success and product profitability [5]. There are at list two main parts of ergonomics: overall physical load and local muscular load.

Measurement methods based on spending muscle forces are used especially in Czech and Slovak Republic. Other countries measure local muscular load more subjectively, through evaluation sheets (ergonomic checklists). However, methods presented in this article are more precise and objective.

In Czech and Slovak Republics, the EMG measurement is a standard method of diagnosing the carpal tunnel syndrome. When doing medical examination of the nerve through the carpal tunnel, a lower sensitivity as well as prolonged distal motor latency occurs. Moreover, a dispersion of potentials is often revealed. If suffering from severe disorders, a lower number of amplitudes of a summing operational potential is typical. According to the values measured, we quantify and stratify the seriousness of disorder in the field of neurography. In the Czech Republic, there is the Standard of Electrophysiological Carpal Tunnel Syndrome Diagnosis effective for the purposes of reporting on occupational diseases [6].

The research team focused on the illnesses caused by local muscular load because they are the most frequent occupational diseases in the Czech Republic. Among them, occupational carpal tunnel syndrome (CTS) is the most frequent. Minks et al. [6]. note that "Carpal tunnel syndrome is the most common mononeuropathy and the most common occupational disease [6]. Also the costing point of view it is very important in the context of healthcare [7]. For elimination of LML we can use the ABC costing. An activity based costing system assigns costs as they actually exist at a point in time, not as they should or could be performed [8].

2.1 Integrated Electromyography as a Classical Measuring Method

After identifying the risk factors in the workplace, an evaluation (objectification) followed. The evaluation was carried out with the use of integrated electromyography (EMG) to generate standardized measurements. In the Czech Republic, the EMG method is the only officially recognized method for measuring local muscular load, and only the authorized companies can carry out the measurements.

The expended muscle strength, the number of movements and the operating position of the limbs are identified and verified during the evaluation of local muscular load, depending on the extent of the static and dynamic parts of a person´s work during an average eight-hour shift.

The EMG method has been widely used in the past to investigate central fatigue, by evaluating gears responses of the central nervous system (i.e. changes in the motor unit recruitment). The methods for measuring can vary a little, especially during prolonged maximal voluntary contraction (MVC), whereby the EMG amplitude decreases progressively in parallel with the force output [9, 10, 11]. Conversely, during submaximal contractions, whereby the EMG amplitude progressively increases in order to retain the force output [12, 13, 14]. Other methods exist around the world measuring muscular activity, e.g. surface electromyography (sEMG). A portable EMG system (MyoGuard) was used to collect myoelectrical signals on-line and analyze them using a computer [15].

Some articles confirm that it is possible to decrease local muscular load in factories by using mini-load multi-shuttle Automated Storage and Retrieval Systems (AS/RS) and discrete event simulation. The simulation model for multi-shuttle AS/RS consists of two lines (single depth) of Storage Racks (SR), an SR machine, Input and Output locations (I/O Location), and other manipulation equipment [16].

However, for typical measurements of local muscular load in the Czech Republic (to obtain the values and type of expended muscular strength of extensors and flexors of the forearm of both upper limbs of a measured person), non-invasive electromyography was carried out using a Holter EMG. The Holter EMG (i.e. data collection station) is shown in Figure 1.

Figure 1
Holter EMG [17]

As Figure 2 shows, five electrodes were placed on both upper limbs to scan the muscle activity. Two electrodes were located on the extensors, two on the flexors and one ground electrode on the tendon in the elbow.

Figure 2
EMG electrodes placed on the arm (own processing)

Prior to taking measurements, it was necessary to individually determine the maximum muscle strength for each upper limb. Each limb was put into a predetermined position (upper arm parallel to the body, with the forearm held at a right angle). The highest activity of electrical potential was recorded as 100% Fmax for measuring the muscle groups of the flexors and extensors of the forearm.

2.2 The Datalogger

This invention is an ergonomic device monitoring local muscular strain. It can be used at workplaces with a higher muscle strain, mainly the one of arms when having frequent/repetitive motions often connected with higher muscle strength. These are particularly workplaces of assembly, but also of other types as well.

Within this project, so-called Measuring device consisting of the Datalogger and measuring gloves was designed and tested. Altogether, four functional samples of Datalogger and measuring gloves were made. The practical use can be seen in Figure 3.

Then, a methodology to measure local muscular strain of employees of manufacturing as well as non-manufacturing companies was suggested. These strains are mainly caused by tunnel syndromes, e.g. carpal tunnel syndrome. A follow-up program application that forms an integral part of the methodology was prepared. This application enables users of the measuring device (Datalogger and measuring gloves) to have a simple and clear measurement output, in the form of graph as well, without having deep knowledge of statistical methods.

The development of this device took place at workplaces of Tomas Bata University in Zlín and Department of Cybernetics and Biomedical Engineering at VŠB – Technical University of Ostrava following standard procedures of HW development in the following stages.

After identifying the risk factors in the workplace, an evaluation (objectification) followed. The evaluation was carried out with the use of integrated electromyography (EMG) to generate standardized measurements.

Figure 3
Sensors, their placement and course of measuring (own processing)

3 New Method of Measuring

3.1 Major Problems the 1st Generation of the Datalogger and Their Solutions

The following major problems occurred when measuring by the datalogger device (1st generation):

1) Incorrect data record on SD card – all sequences missing on the data record (many minutes)
2) Failures – sensors FSR 400 and FSR 402 breakages
3) Insufficient length of testing
4) Low weight of manipulated components of the Minnesota test
5) Having problems when doing an automatic reading of number of motions from the figures of gyroscope and accelerometer
6) Having problems when using gloves

How we coped with the problems when working on modification of measurement methods and the datalogger (2nd generation) itself:

Ad 1) For the purpose of records, it is necessary to use SD cards with the capacity of maximum 32 GB (SDHC) only. The recommended capacity of SD card is 8 GB. SD cards with higher capacity can have different features and time constants as well as SW of keeping the records; therefore, can cause loss of records from the datalogger, i.e. not saving the data.

In the course of pilot measurements, precautions to eliminate the problems with keeping the records on SD card were taken as well as hardware tuning up of the

datalogger was done in order not to have an incomplete data record from the sensors. Neither the influence of battery power nor the source of power supply (batteries vs. transformer) was confirmed when having problems with records.

Figure 4
Datalogger - apparatus (own processing)

Ad 2) The sensor breakage is a matter of mechanical wear of these items when testing (taking measurements), when in contact with fingers and the tool a proband is working with. A very simple and efficient protection showed up to be putting a thin adhesive tape on a sensor. The sensor will be protected against the breakages and its lifespan will be longer by using this simple method. Moreover, the measurement results are not distorted at all. For this reason, the number of measurements done by using a particular sensor is approximately 20-50 depending on the type of grasp, work operation, temperature and other aspects of measurements. In other words, having a lifespan of each individual sensor for 100 measurements is possible in exceptional cases only, i.e. under laboratory conditions.

Ad 3) Approx. five-minute length of the test, which is an average length of the standardized Minnesota test, showed up not to be sufficient from time perspective for statistical evaluation. This is for the reason that it was necessary not to consider and evaluate the first sequence of measurement when a proband was getting "initial training".

We have to consider this when taking measurements or doing analysis of real work in practice as even an experienced worker can be put out of countenance (doing usual routine work), e.g. an attached sensor to his/her finger. It is also the feeling that he/she is being monitored, evaluated, filmed, etc.

Therefore, it is necessary to include the 1st part of measurements, but the phase of training must be read of the total process, and the measurement itself should take

approx. 20 min. This is optimal regarding the record and evaluation frequency of 50-60 seconds. From mathematics perspective, it is for the reason that the interval of 50-60 seconds, when being statistically evaluated (see Chapter 7.4. The evaluation of correlation coefficient dependency on a sample frequency for sample measurements), seems to be the most convenient for the rate of final record correlations (values of approx. 0.9).

Moreover, it provides 1,200 records for calculations and the analysis of particular work (using the record after 60s; 20x60s in total).

Ad 4) As the grasp forces were not sufficient for the reason of low weight of manipulated components of the Minnesota test, the chronic fatigue syndrome which results is the carpal tunnel did not occur. Therefore, the components of the Minnesota test as well as the test itself was substituted by a model test, the so-called "puck test". This test is of a similar type; only different numbers of motions and heavier manipulated objects (340g) are considered.

Ad 5) During the pilot measurement itself, it was evident that it is not possible to do reading of number of motions from the gyroscope and accelerometer as the automated motion and trajectory motion frequencies evaluation requires a 3D scan of the rooms as well. Also, the sensor attached to the wrist (gyro + accelerometer) enables us to measure "the angle of the wrist only" (wrist x forearm), but not the work done by fingers in the constant "angle of the wrist".

Ad 6) Using thin gloves (made of nitrile) showed up to be without any problems under laboratory conditions. Nevertheless, further check-up measurements will have to be taken for thicker gloves.

3.2 Measurement Methods

In this subchapter, a more accurate description of checkpoints for practical measurements will be presented. Five methods systematically divided among three checkpoints were used. In the first one, training and filling in the questionnaire was done. In the second checkpoint, the Fmax figures were recorded, EMG and the measuring device were placed, and the load test was done. The last checkpoint was BIA (Bio Impendence Analysis).

1) BIA a certified device
2) EMGH (EMG Holter Measurement); a certified device
3) Tension meter measurement (tension meter – a set)
4) Unique non-standardized load test
5) Measuring device (developed by Department of Cybernetics and Biomedical Engineering at VŠB – Technical University of Ostrava)

3.3 Checkpoint – Questionnaire Filling in and BIA

The probands went through a proper training before starting taking measurements. The goal of a project was explained to them, the whole course of measurements was introduced including the methods used, and consequently, a trainee task was described. First, they had a chance to see it on video recording, and then, they could practice is by themselves. The purpose of this training was to eliminate mistakes during the measurement itself. Moreover, we expect that this could have minimized certain rate of nervousness causing inaccuracy and a lower coordination during the measurement.

In the first checkpoint, the questionnaire of a proband was filled in with the aim to record his/her identification data, laterality and mainly his/her general state of health. A part of it was BIA done by InBody230 device, which was operated by a specialist.

3.4 Checkpoint - EMG

This checkpoint included a tension meter as well as device recording electric muscle activity, the so-called EMGH = electromyography holter. A maximum fist clench was measured – the elbow joint in the angle of 90°.

The EMG device was used before measuring a maximum fist clench by a tension meter (Fmax.; measured in Newton units, N). This was done lengthwise following two muscles; one of them located at the dorsal part of a forearm and the second one on the ventral part. The electrodes were attached to a muscle pad by using the adhesive tape; in the electrode, an electrode cream was applied for optimization of a signal transmission.

Moreover, every electrode itself was secured by another adhesive tape. Consequently, the EMG recording started in order to measure a maximum muscle activity when measuring F_{max}. Each of the probands was measured by a tension meter **twice** for each upper limb. A higher value measured out of two attempts was recorded (measured in Newton units). This was always done in the same order; first, right upper limb (RUL) measurements and then, left upper limb measurements (LUL).

4 Results - Statistical Data Processing

4.1 Unprocessed Datalogger Data Processing

The datalogger data recorded on SD card are then divided on a personal computer by using EGParser application into 7 files. The analogue data, keeping record of AD non-dimensional values, are part of one file only.

Processing, the datalogger and EMG holter evaluation of values and the dependency analysis are done in MS Excel. The information regarding muscle forces used in the course of measurement is presented in the form of analogue data one after another in six columns for LUL (left upper limb) and six columns for RUL (right upper limb) for every individual sensor. Every record starts with a time code – in the first column, consisting of a number defined by seconds; in the following column of an order number in particular second.

The statistical data processing from the datalogger starts by a thorough check-up of the records kept throughout the whole period of measurements. As standard, the device keeps record of between 99% to 100% of data, which represents a record loss of 0-1000 lines (0-10 seconds) per 100 thousand lines of a record.

Consequently, the noise of every individual sensor is filtered out. For the reason of pressure of the sensor on every finger by using the adhesive tape, "a minimum load" is set. This level (minimum values needed for the sensors to start measuring) is then, in the course of statistical processing, taken away from all values of a particular column. The same procedure is used for all sensors attached, or more precisely the data columns in the output files for LUL and RUL.

Then, millisecond records of AD values are re-calculated by using the equation for calculation of force in Newton units (see Chapter 3.2), and the forces for each upper limb are add up. Due to this, two input variables for left and right upper limb are obtained, i.e. a record of real forces used during the task.

By averaging, the values regarding a muscle activity per second are obtained in order to analyze dependency in relation to the EMG record of values measured on-line per second by using PC. Moreover, the analysis keeps track of correlation regarding one-second, ten-second and sixty-second average values of both devices. A progressive increase in correlation coefficient (sixty-second average values – most often between 0.7 and 0.9) confirms conformity with the primary research hypothesis.

4.2 Research Reliability and Validity

Both devices measure muscle load when making simple mechanical changes in muscle activities. The force used by the person tested is a metric variable which varies in the course of testing – different levels of location and variability – providing consistent information in certain time sequence.

J. Hendl: "The reliability is expressed by a relative reliability coefficient, and we study it by using several methods, e.g. parallel test measuring, which means a measurement conformity with other adequate measurement of the same construct." [18]. In our procedure, a maximum reliability is offered thanks to using the EMG holter, which is then studied by algorithm of searching for the highest conformity – correlation.

The correlation coefficient as a primary correlation measurement indicator for both devices.

The expected linear dependency between the EMG values and the muscle forces used monitored by the sensors is studied by using statistical analysis methods, to be more precise the correlation and regression analyses. For the newly developed measuring device values, a strong dependency is expected. By using a regression model, the dependency with a high rate of explained variability and the incorrect component E is revealed. Moreover, an extreme systematic deviation occurred.

The measurement validity is an ability to clarify what the object of measuring is (offering correct information, which means that "the measuring procedure really measures what we expect to be measured"). According to Hendl [18], the validity concerns "adequacy, meaningfulness and utility of specific conclusions that are based on measurement results". The validation of measuring methods is a process supporting such belief. The validity confirmed under certain circumstances (context + purpose) is valid for this context and identical purpose only [19]. When checking-up validity, the content validity, criterion and construct validity are studied [18].

4.2.1 Content Validity

We check to what extent our measurement really represents typical features or the quality [18]. In order to have the content validity of good quality, it is essential to specify the area of measuring itself [20]. Regarding the content validity, our measurement really represents the feature, i.e. it tests a minor muscle load of a hand and forearm when manipulating. The newly developed device (the datalogger) can measure the muscle force used on fingers and the palm. An average muscle force for grasp vs. time is a quantity precisely evaluating the degree of a muscle load which a tested subject must cope with while working. It includes an aspect of fatigue which is really important for us when studying occupational diseases.

4.2.2 Criterion Validity

We assess results and their consensus of the newly developed procedure with another type of measurement that has been verified earlier [18]. When assessing the criterion validity, the results obtained can be compared to other values characteristic for a particular criterion. In our measurements, we wanted to compare the results from our device with the results measured by using the

Minnesota manipulative tests that are standardized with having the norms processed. The aim was to prove that both tests can reveal the same failure in performance, i.e. people performing above standard according to the Minnesota test will be detected by our measuring device and the parameters measured as well. In our country, the EMG measurement is a standard method of diagnosing the carpal tunnel syndrome. When doing medical examination of the nerve through the carpal tunnel, lower sensitivity as well as prolonged distal motor latency occurs. Moreover, a dispersion of potentials is often revealed. If suffering from severe disorders, a lower number of amplitudes of a summing operational potential is typical. According to the values measured, we quantify and stratify the seriousness of disorder in the field of neurography. In the Czech Republic, there is the Standard of Electrophysiological Carpal Tunnel Syndrome Diagnosis effective for the purposes of reporting on occupational diseases [6]. We checked whether the EMG results will be identical to the ones measured by the datalogger. This should be a valid method of finding out about the nerve conductivity disorder (in this case of a central nerve) as this may influence the force of innervated muscles (i.e. wrist flexors and 1^{st}-3^{rd} finger). The method of studying the criterion validity most often used is the correlation coefficient calculation between these two values.

4.2.3 Construct Validity

It concentrates on theoretical aspects of the construct measured (variables, quantities). The evidence of the construct validity is of convergent or discrimination character. If it is of convergent character, the test confirms relations to the variables that we expect from theory. If it is of discrimination character, there is no relation of the test to the variables when we do not expect the relation. In this phase of testing a new procedure, it is important that the results are the ones expected from theory [18].

To assess the construct validity, there is a need for evidence that the test really measures the feature or characteristic needed. We check whether the datalogger measurements of a minor muscle load of workers in practice can reveal a lower performance of those who have suffered from the carpal tunnel syndrome or other occupational diseases. The content validity focuses on measuring the feature researched as a whole. It is researched by studying the literature, theory of the feature, seeing the previous empirical research and the expert opinion. The third one from a group of tools for identifying validity is, on the contrary, the criterion validity. This one is well-measurable by comparing it with certain objective and already validated criterion. This is done either concurrently, the so-called concurrent validity, or prospectively, the so-called predictive validity. The last approach to identify validity is the construct validity. This one is proved when a mutual relationship between the indicator and other variables is found out from the data obtained and which we would expect from studying theory. The convergent validity is proved when one indicator of the concept researched is

associated with other indicators of the same concept, but of different character. On the contrary, the discrimination validity is proved when the variable values measuring the concept are not strongly associated with measuring different, but related concepts. The basic tools for measuring the convergent as well as discrimination validity are the exploratory and confirmatory analyses and their various aspects. From theoretical and methodological perspective, the MIMIC model is the most complicated approach to measuring the construct validity. We used this theory also in some other researches [21].

This one places a measuring tool of research, usually a range of these, in the context of conceptually similar as well as different variables. In addition, it specifies possible causal relations between variables, and compares the data with theory. In case these correspond with the theory, this finding confirms the construct validity of the variable studied.

5 Discussion

Five key probands were chosen for detailed analyses when all well-known problems had been checked carefully for 20 minutes. There were also the following precautions taken: the frequency setting for turning the pucks according to metronome, the sensors being attached to four fingers not to be taken off, keeping a double-record of values – on SD card of the datalogger thanks to a particular format and type of the SD card, a record check-up throughout the whole measurement period, a check-up of all sensors and their functionality right after the end of measurements.

These evaluations were carried out without any support of a later-developed SW application**.** The aim was to find dependency between the EMG measurements and the muscle force recorded by the datalogger. The value averaging set per every minute was later confirmed to be very close to an optimal sampling frequency. As an illustration, the records of two probands are mentioned below. But in total we tested more than 100 probands and for about 30 operators at the production lines (Figure 6). The rest of elevations, in a form of unprocessed data, were added on the CD-ROM enclosed (together with the application developed). All this data were handed to the Moravian-Silesian Automotive Cluster c.a. (hereinafter referred to as MAK) and Ministry of Industry and trade Czech Republic and (MIT CR), and to The Industrial Property Office in Czech Republic (IPO CZ) too.

Table 1
Dependency between the EMG measurements (Flexors LUL,LUR) and the muscle force recorded by the datalogger (LUL, RUL datalogger): Proband A – 1x 20 min

Time	Flexors LUL	Flexors RUL	LUL-datalogger	RUL-datalogger
12:00:17	14.50	12.18	3.49	2.75
12:01:17	13.47	11.50	3.2261525	2.8365554
12:02:17	13.18	10.01	3.3788313	2.4720185
12:03:17	13.47	10.46	3.7328694	2.6778192
12:04:17	11.06	9.39	3.2591333	2.7932303
12:05:17	10.95	9.16	2.4964413	2.1003172
12:06:17	10.89	9.12	2.6344152	2.1253838
12:07:17	9.85	9.05	2.8381982	1.9965359
12:08:17	11.25	9.20	3.0447249	2.2497033
12:09:17	10.68	8.95	2.9297469	2.1580144
12:10:17	10.62	9.58	2.73	2.31
12:11:17	10.75	9.62	2.9933314	2.4385115
12:12:17	9.79	9.28	2.6689625	2.3447799
12:13:17	9.12	8.18	2.5253218	2.0219681
12:14:17	10.17	7.90	2.641671	1.936453
12:15:17	10.02	8.48	2.6497252	2.0183217
12:16:17	10.41	9.01	2.5791602	1.9634366
12:17:17	9.60	8.61	2.5194543	1.9121674
12:18:17	9.38	9.02	2.4235038	1.9146724
12:19:17	9.78	9.84	2.7883636	2.0028123
CORRELATION	LUL	RUL		
	0.86243	0.79174		

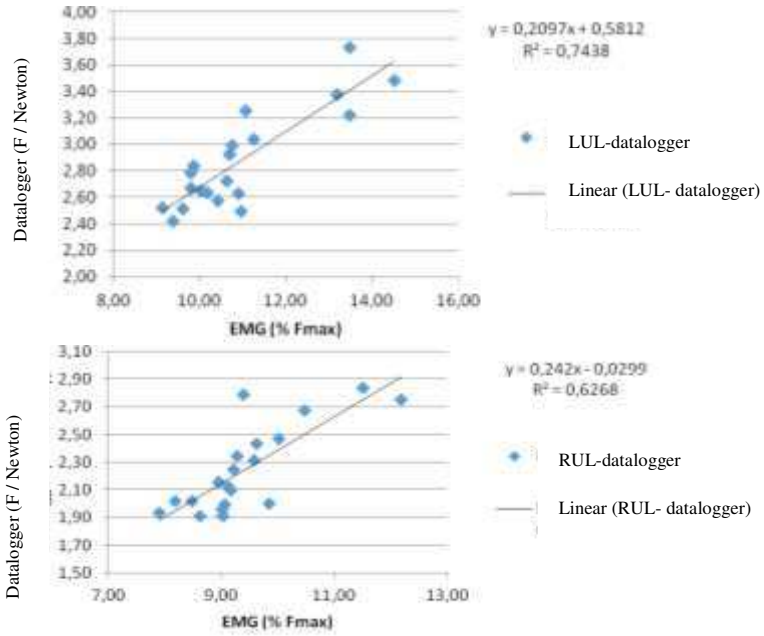


Figure 5
The demonstration of evaluation of A proband´s left (up) and right upper limb (down) (own processing)

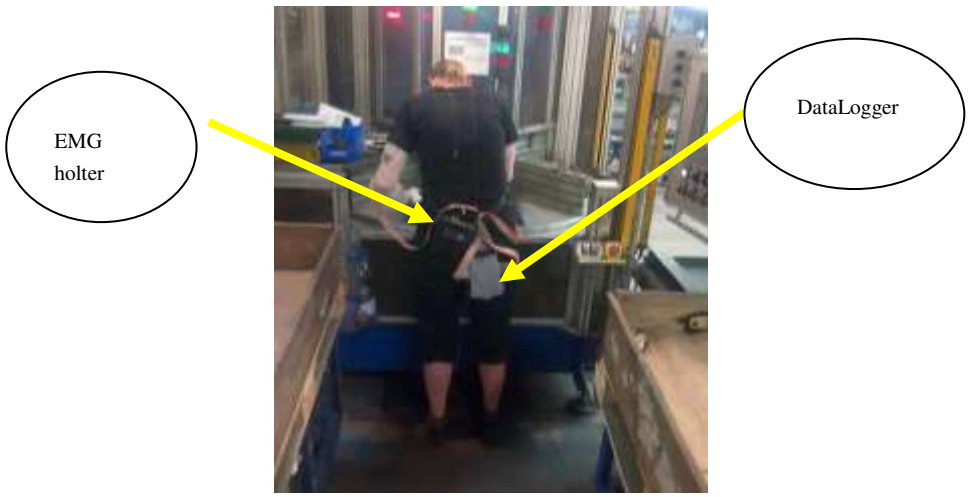


Figure 6
Measuring in factory: dependency between the EMG measurements and the muscle force recorded by datalogger: Operator in ITT holding Ostrava (own processing)

5.1 The Dependency Evaluation of Correlation Coefficient and the Sampling Frequency for Sample Measurements

In this chapter, the last illustration of the final data required for the input elevations is presented. The input data is the sampling frequency of 1 second (the EMG holter implicit configuration and the datalogger adjusted means). Then, the correlation analysis is done depending on the alteration of this sampling frequency (see Figure 7). In the Figure 8, the correlation coefficient dependencies on the sampling frequency (in seconds) are illustrated.

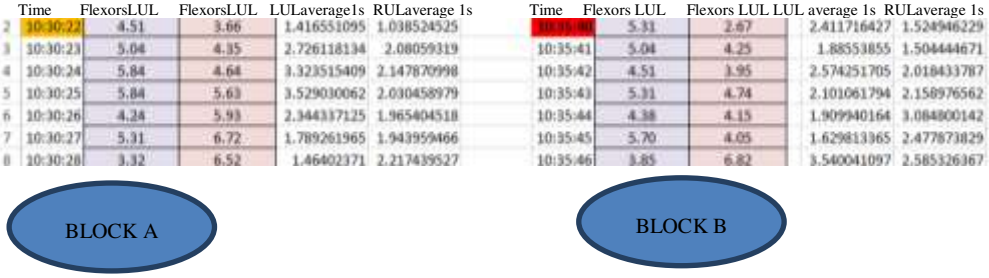


Figure 7

The illustration of the input data from MS Excel for further processing in R software (own processing)

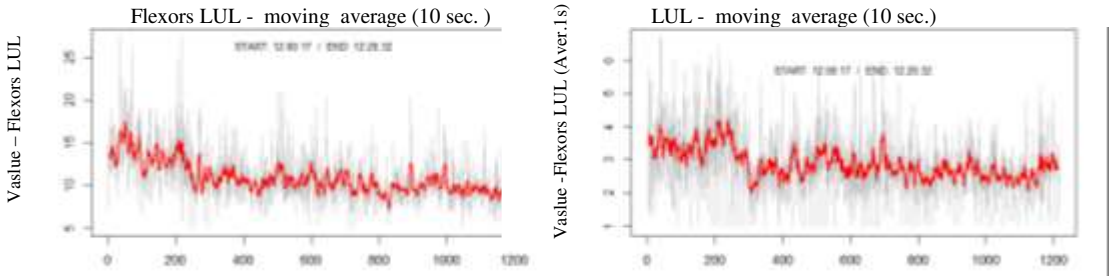


Figure 8

The course of testing – 1st check-up – a graphical debugging of autocorrelation enables to find significant deviations (own processing)

The final analysis output is a graphical comparison of average muscle forces (the datalogger) and EMG values in accordance with legislation which sets occupational health rules.

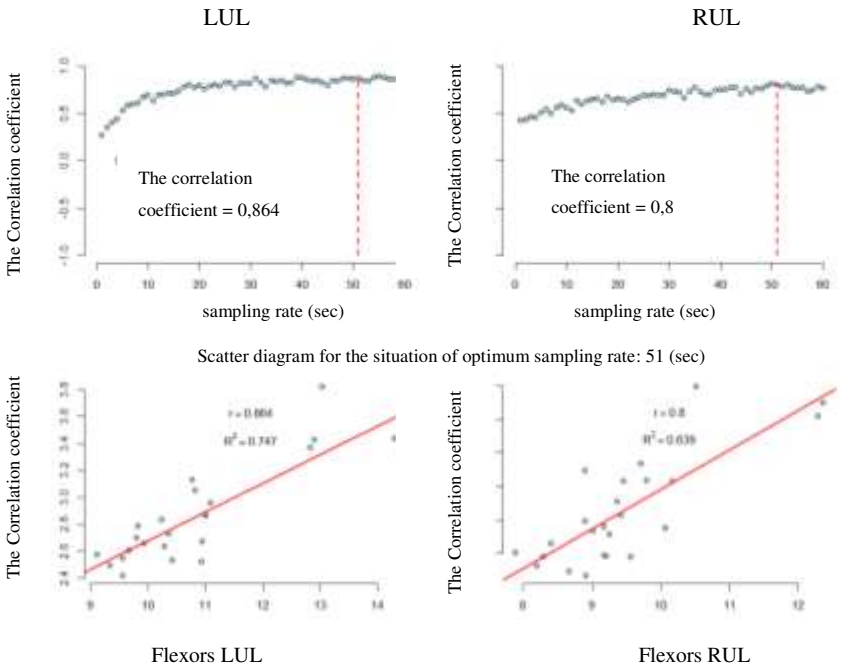


Figure 9
Finding an optimal sampling frequency; based on this, the dependency as being a linear regressive model is illustrated (own processing)

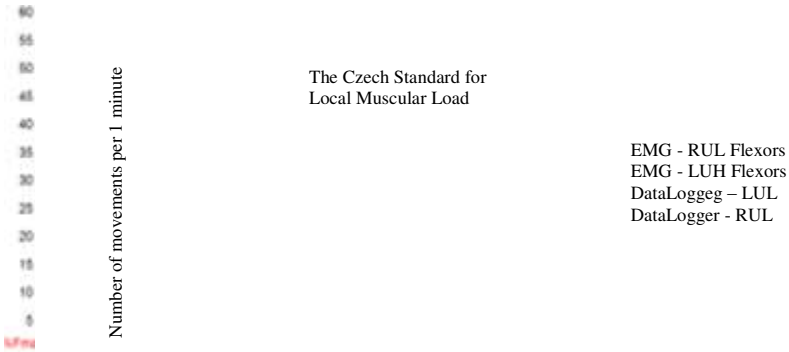


Figure 10
The data output of both devices evaluating a local muscular load for both upper limbs (own processing)

The Figure 10 shows that the EMG results confirm the work to be hard – the values exceeding the norm and breaking occupational health rules. The datalogger as well as the EMG device confirmed a higher muscle load of the right upper

limb. For the reason of no similar limits for the load itself, the values can be compared visually only.

Conclusions

This datalogger was patented by The Industrial Property Office in the Czech Republic (IPO CZ) in 2017, under the group patent / application number PV / 2015-820 and it holds the security document code 306 627. IPI CZ is a central body of state administration of the Czech Republic acting in the field of industrial property protection. It is necessary define the limits of this new apparatus and methodology because this apparatus was patented only by The Industrial Property Office in Czech Republic. It means that its application is possible mainly in the Czech Republic and Slovak Republic, where the legislations are the same. Despite the high level of dependence resulting values measured (see Figures 7-10 and Table 1) by both devices (EMG vs. logger) the datalogger has no certification by The National Institute of Public Health (NIPH) is a health care establishment for basic preventive disciplines - hygiene, epidemiology, microbiology and occupational medicine of course in the ergonomics context as well.

The possibilities of a new datalogger were presented, mostly in laboratory conditions. However, this apparatus was tested at dozens production lines and workplaces as well. Our results have confirmed, that we can measure the load of all fingers individually. The load in the Table 1 (columns: LUL-datalogger, RUL-datalogger) is the sum of the load of all the fingers of the hand measured. The advantages of the new measuring principle are: more accurate measurements, higher frequency data reading, video analysis connected with the impulses of gloves, to scan the force directly at the tip of every finger excluding the effect of the operator distortion.

Acknowledgements

The authors are thankful to the Internal Grant Agency of FaME TBU No. RO/2016/03, VaV-IP-RO/2016/03 (“Parameters modelling for effective production and administrative processes in industrial companies based on Industry 4.0 concept”) and Internal Grant Agency of FaME TBU No. IGA/FaME/2015/030 (Evaluation model of ergonomic principles in Czech companies) for the financial support to carry out this research.

References

[1] Edl, M., Lerher, T., Rosi, B. (2013): Energy efficiency model for the mini-load automated storage and retrieval systems. International Journal of Advanced Manufacturing Technology. 2013: 1-19, ISSN: 0268-3768

[2] Belás, J., Bartoš, P., Habánik, J., Novák, P. (2014): Significant Attributes of the Business Environment in Small and Meduim-Sized Enterprises. Economics and Sociology, 7(3):22-39, DOI: 10.14254/2071-789X.2014/7-3/2

[3] Definition and Domains of ergonomics (2015): International Ergonomics Association (IEA) [online]. [cit. 2015-11-18] Available from: http://www.iea.cc/whats/index.html

[4] Rowan, M., Phillip, P., Wright C. (1995): Ergonomics is good for business. Facilities. (13) 8: 18-25. DOI: 10.1108/02632779510090810. ISSN 0263-2772

[5] K. Fisher, C. Hobelsberger, K. J. Zink (2009): Human Factors and Sustainable Development in Global Value Creation. In: IEA (Ed.): 17th World Congress on Ergonomics of IEA 2009. Bejing. IEA Press

[6] Minks, E., Minksová, A., Brhel, P. and Babicová, V. (2014): Profesionální syndrom karpálního tunelu. Neurologie pro praxi [online] 15(5) Available from: http://www.neurologiepropraxi.cz

[7] Popesko, B., Tučková, Z. (2012): Utilization of process oriented costing systems in healthcare organizations. International Journal of Mathematical Models and Methods in Applied Sciences. 6 (1): 200-208

[8] Rajnoha, R., Chromjaková, F. (2009): Activity-based costing and efficiency of its application in the wooden houses production. DREWNO-WOOD, Poznaň : Instytut Technologii Drewna, Poland. 52 (181):105-127, ISSN 1644-3985, 2009

[9] Bigland-Ritchie, B., Johanson R, Lippold, CJ, Woods, JJ (1983): Contractile speed and EMG changes during fatigue of sustained maximal voluntary contractions. J. Neorphysiol. 50 (1): 313-324

[10] Sirin, AV., Patla AE (1987): Myoelectric changes in the triceps surae muscles under sustained contractions. Eur J Appl Physiol. 56: 238-244

[11] Hatzikotoulas, K, Siatras, K, Spyropoulou T, Paraschos, E, Patikas, D (2004): Muscle fatigue and electromyographic changes are not different in women and men matched for strength. European journal of applied physiology. 92 (3): 298 -304, DOI 10.1007/s00421-004-1095-4. ISBN 1439-6319

[12] Fuglevand, AJ., Zackowski, KM., Huey, KA., Enoka, RM. (1993): Impairment of neuromuscular propagation during human fatiguing contractions at submaximal forces. J Physiol (Lond). 460: 549-572

[13] Gamet, G., Maton, B. (1989): The fatigability of two agonistic muscles in human isometric voluntary submaximal contraction:an MG study: I. Assessment of muscular fatigue by means of surface EMG. Eur J Appl Physiol. 58: 361-368

[14] Lascher, WN., Cresswell, AG., Thorstensson, A. (1996): Central fatigue during a long lasting submaximal contraction of the triceps surae. Exp Brain Res. 108: 305-314

[15] Rolander, B., Jonker D., Karsina, A., Öberg, T. (2005): Evaluation of muscular activity, local muscular fatigue, and muscular rest patterns among dentists. Acta Odontologica Scandinavica. 63 (4) pp. 189-195, DOI:10.1080/00016350510019964

[16] Lerher, T., Sraml, M., Potrc, I. (2011): Simulation analysis of mini-load multi-shuttle automated storage and retrieval systems. International Journal of Advanced Manufacturing Technology. 54: 337-348, DOI 10.1007/s00170-010-2916

[17] Gadourek, P, Lebeda, T. (2010): Ergonomická laborator: Mereni a hodnoceni lokalni svalove zateze metodou integrovane elektromyografie. GETA Centrum s.r.o. [online]. GETA Centrum s.r.o. Available from: http://www.plm.automation.siemens.com/cz_cz/Images/Gadourek_a_tcm841-117307.pdf

[18] Hendl, J. (2006): Prehled statistických metod zpracování dat. Analýza a metaanalýza dat. Praha: Portál. ISBN 80-7367-123-9

[19] Disman, M. (2005): Jak se vyrábí sociologická znalost. 3. vyd. Praha: Univerzita Karlova v Praze. ISBN 80-246-0139-7

[20] Trebuňa, P., Halčinová, J., Fiľo, M., Markovič, J. (2014): The importance of normalization and standardization in the process of clustering. SAMI 2014: IEEE 12th International Symposium on Applied Machine Intelligence and Informatics, pp. 381-385, ISBN 978-1-4799-3441-6

[21] Tuček, D. (2016): New strategy for Business Process Management – quantitative research in Czech Republic. International Advances in Economic Research. ISSN: 1573-966X (electronic version), ISSN: 1083-0898 (print version), DOI: 10.1007/s11294-016-9613-9

2017 Reviewers

Babic, Frantisek
Babik, Marian
Baczynski, Michal
Badacsonyi, Ferenc
Balasubramaniam, Jayaram
Barile, Claudia
Bartak, Roman
Batyrshin, Ildar
Bednar, Peter
Békési, Bertold
Bitay, Enikő
Bock, Igor
Bonet, Isis
Borkotokey, Surajit
Borri, Alessandro
Bustince, Humberto
Castro-Espinoza, Félix
Cedomir, Milosavljevic
Čertický, Martin
Chen, Xuan
Chikovani, Valeriy
Copot, Cosmin
Copot, Dana
Csink, László
Csiszárik-Kocsir, Ágnes
Danaj, Adela
Djellali, Choukri
Dombi, József
Dragan, Antic
Drégelyi-Kiss, Ágota
Drexler, Dániel András
Dulf, Eva-Henrietta
Duma, Florin
Eigner, György
Einspieler, Sascha
Elmenreich, Wilfried
Esztergár-Kiss, Domokos
Fam, Kim
Ferenci, Tamás
Fleiner, Rita
Fodor, Bertalan
Fodor, Mónika
Fullér, Róbert
Furdik, Karol
Galántai, Aurél
Gausz, Tamás
Gergely, Kristóf
Glova, Jozef
Gonda, Viktor
Gonda, Zsuzsanna
Gonzalez-Mendoza, Miguel
Goulão, Miguel
Gönczy, Tamás
Gul, Muhammet
Haase, Jan
Halász, Sándor
Harmati, István
Havlice, Zdenek
Horváth, Richárd
Huszka, Péter
Hýrošová, Tatiana
Iclanzan, David
Inenaga, Shunsuke
Ionescu, Clara Mihaela
Jayaram, Balasubramaniam
Kádár, Péter
Katanyukuk, Tatpong
Keppler, István
Kertész, Gábor
Kiss, Dániel
Klešč, Marián
Klespitz, József

Kopacek, Peter
Kovács, György
Kovács-Coskun, Tünde
Krallics, György
Lacko, Peter
Laclavík, Michal
Lakatos, István
Le Capitaine, Hoel
Lovas, Robert
Lukacs, Zsolt
Mach, Marian
Machova, Renata
Mankovits, Tamás
Mathew, Sunil
Mesiar, Radko
Mikó, Balázs
Milosavljevic, Cedomir
Miranda-Jiménez, Sabino
Molnár, Ildikó
Mura, László
Muresan, Cristina
Nagy, Benedek
Nagy, Péter
Nagy, Viktor
Nikolic, Sasa
Nowotyńska, Irena
Oldal, István
Olteanu, Alexandru-Liviu
Ondas, Stanislav
Pai, Tun-Wen
Palúch, Stanislav
Paralic, Jan
Pavlovičová, Jarmila
Peric, Stanisa
Pethes, Róbert
Pichardo-Lagunas, Obdulia
Pinto, Carla
Piros, Péter
Pokorádi, László
Ponce, Hiram
Posadas-Duran, Juan-Pablo
Qiu, Jianbin
Rea, Pierluigi
Riha, Zdenek
Rychtáriková, Monika
Sápi, Johanna
Sarkoci, Peter
Sarnovsky, Martin
Saso, Blazic
Schiaffino, Silvia
Sedlák, Vladimír
Sekaj, Ivan
Sevcovic, Daniel
Sidorov, Grigori
Šimko, Jakub
Šipošová, Alexandra
Siran, Jozef
Spirkova, Jana
Stupnanova, Andrea
Svátek, Vojtěch
Szabó, László Zsolt
Szabó, Tamás
Szabolcsi, Robert
Szénási, Sándor
Takács, Árpád
Takács, Márta
Tar, József
Tomasek, Martin
Új, József
Vajda, István
Vajk, István
Vámossy, Zoltán
Varga, Árpád
Varga, János
Varga, Péter János
Várlaki, Péter
Vavríková, Lucia
Vemuri, Nageswara Rao

Vranić, Valentino
Wang, Xi
Wang, Yong
Zaharia, Sebastian Marian
Zamuda, Ales
Zaytsev, Vadim
Zichar, Marianna
Zolotova, Iveta