Improving Construction Procurement Systems using Organizational Strategies

Heap-Yih Chong

Faculty Engineering & Science, Universiti Tunku Abdul Rahman, Setapak 53300, Kuala Lumpur, Malaysia
e-mail: chonghy@utar.edu.my

Christopher N. Preece

Razak School of Engineering & Advanced Technology, Universiti Teknologi Malaysia, 54100, Kuala Lumpur, Malaysia
e-mail: chrispreece@ic.utm.my

Abstract: Numerous types of construction procurement systems have been developed for project implementation. However, previous studies have not focused on subsequent managerial strategies and the project organizational forms to be adopted towards the selected procurement system. This research proposes that further managerial theories are required to enhance the project performance and effectiveness. Therefore, this paper aims to extract the principles of projectized and nonprojectized organizations and incorporate them with the selected procurement systems at the project level. A mechanism for assessing the key areas of compatibility was developed using the well established McKinsey 7 S model. The paper shows that the characteristics of the organizational principles are complementary with the procurement systems. It contributes an insight for future strategic organization and management at the project level in construction.

Keywords: projectized; nonprojectized; construction procurement systems; organizational principles; McKinsey; compatibility

1 Introduction

The construction industry is a highly fragmented sector due to the conventional practices [17] and supply chains [8]. This fragmentation inhibits the performance and effectiveness of construction projects. Well-coordinated organizational strategies are required to enhance communication and provide a clear understanding of the relationships and interdependencies among the stakeholders involved [4, 24, 35].

At the outset of construction projects, the managerial strategies include the selection of appropriate procurement systems or building-delivery methods to secure the success of the project [6, 7, 9, 21, 28, 29]. In addition to traditional, design and build systems, innovative procurement systems have been developed such as improved private finance initiatives; public–private partnerships [13]; and specialist task organizations [36].

Today, many theories have been developed to address the emergence of increasingly complex organizational forms [18]. Yet, none has investigated on the selected procurement systems in terms of project management strategies and organizational forms. Hence, this paper aims to address this gap in knowledge by focusing on how project-based organizations, namely, projectized or nonprojectized organizations might be compatible with selected procurement systems for the successful implementation of projects. The review focuses on the micro level of project-based organizations, as opposed to the conventional approach, which is looking at the macro level of an entire organization [1]. The micro level incorporates the principles of project management strategy and organization into a single-project base. It was decided to focus on the lump-sum and design-and-build procurement systems for the purpose of this study.

It is believed that this review would add value to the existing project-management body of knowledge (PMBOK) in construction by examining how these construction-procurement systems could be implemented through projectized and nonprojectized organizations. The purpose of examining the compatibility of these management strategies and organizational approaches is to improve the process of marshalling the adequate human, material, and financial resources [44], as well as to enhance organizational performance and capabilities [15, 22, 43, 46]. Ultimately, it would provide a different perspective to the management organization of procurement systems in support of project effectiveness.

2 Principles of Projectized and Nonprojectized Organizations

Classical management principles consist of planning, organizing, staffing, controlling, and directing. At the macro or project level, organization is defined as a group of persons brought together for a purpose or to perform some type of work within an enterprise (Project Management Institute) [40]. Moreover, Kerzner (2006: 4) has defined project management as "*the planning, organizing, directing, and controlling of company resources for a relatively short-term objective that has been established to complete specific goals and objectives*". Each principle of project management has specific roles and activities to be performed at the micro level of a project, for instance, the organizational breakdown structure is a hierarchically organized depiction of the project's organization arranged so as to

relate the work packages or a breakdown of the skill sets of the people who carry out the work in the various organizational units [40, 44]. The management of a company needs to integrate from the business or company level, to the project level and to develop management strategies and a planned and systematic approach to project organization [2].

Projectized organization describes as any organizational structures in which the project manager has full authority to assign priorities, apply resources, and direct the work of persons assigned to the project [40].

The projectized organization can be further developed as a division within a division, whereby project managers maintain complete line authority over a continuous organizational flow of the project [24]. In other words, staffs report to only one person, which enables a strong communication channel within projects. Furthermore, team members are often collocated [40]. Learning and adoption of innovative solutions must be cultivated within the project's framework for developing its often highly customized outputs [2]. At the project level, the projects in projectized organizations are undertaken within a structure tailored to manage uncertainties or risks [10]. In addition, people require to undergo certain project-specific training and to possess specific competencies in project management.

The principles and theories of projectized organization are relatively new from a micro level perspective. A detailed empirical study needs to be carried out to identify the features of a projectized organization at the project level. Table 1 shows the implementation features that can be applied at the project level, obtained from literature review, which have been mainly modified from the advantages and disadvantages of projectized organization [24].

In the traditional approach, organization theories mainly classify the hierarchy structure for the upper levels of management. The features could be innovated and referred to at the project level. It would increase the project effectiveness and outcomes in terms of a proper coordination between the project manager and staff in the organizational structure.

On the contrary, the terminology of ***nonprojectized organization*** is rather new in organizational principles. Some describe it as a production-based [38] or function-focused [16] organization. It a habitual management process for low-complexity and highly routinized work [10]. Nevertheless, the principles are related or derived from the classical functional organization, the management structure of which has survived for more than two centuries [24].

Table 1
Implementation features of projectized organization in project level

Projectized Organization in General [24]	**Projectized Organization in Project Level**
Advantages • *Provides complete line authority over project* • *Participants work directly for the project manager. Unprofitable product lines are easily identified and can be eliminated* • *Strong communication channels* • *Staffs can maintain expertise on a given project without sharing key personnel* • *Vary rapid reaction time is provided* • *Personnel demonstrate loyalty to the project; better morale with product identification* • *A focal point develops for out-of-company customer relations* • *Flexibility in determining time, cost and performance trade-offs* • *Interface management becomes easier as unit size is decreased* • *Upper-level management maintains more free time for executive decision-making*	Implementation Features • *Full authority in a maximized communication structure for Project manager* • *Direct control to participants and project lines* • *Quick response as easy interface management* • *High loyalty and morale for personnel* • *Flexibility but lack of interchange opportunities between projects* • *Higher cost* • *Poor transferring of knowledge management or perpetuation of technology*
Disadvantages • *Cost of maintaining this form in a multiproduct company would be prohibitive due to duplication of effort, facilities, and personnel; inefficient usage* • *A tendency to retian personnel on a project long after they are needed. Upper-level management must balance workloads as projects start up and are phased out* • *Technology suffers because, without strong functional groups, outlook of the future to improve compnay's capabilities for new programs would be hampered (i.e., no perpetuation of technology)* • *Control of functional (i.e., organizational) specialists requires top-level coordination* • *Lack of opportunities for technical interchange between projects* • *Lack of career continuity and opportunities for project personnel*	

The nonprojectized organization produces identical or similar projects with minimal uncertainties and variations [38]. People working on such projects do not really require substantial project-management knowledge and this type of organization consumes human and material resources much above normal operations [10]. Before the implementation features of nonprojectized organization can be identified, the advantages and disadvantages of a classical organization need to be reviewed. Table 2 lists the advantages and disadvantages of a classical or functional organization.

Table 2

Advantages and disadvantages of classical organization (Kerzner, 2006)

Advantages	**Disadvantages**
• *Easier budgeting and cost control are possible* • *Better technical control is possible (specialist can be grouped to share knowledge and responsibility; personnel can be used on many different projects; and all projects will benefit from the most advanced technology)* • *Flexibility in the use of manpower* • *A broad manpower base to work with* • *Continuity in the functional disciplines; policies, procedures, and lines of responsibilities are easily defined and understandable* • *Admits mass production activities within established specifications* • *Good control over personnel, since each employee has one and only one person to report to* • *Communication channels are vertical and well established* • *Quick reaction capability exists, but may be dependent upon the priorities of the functional managers*	• *No one individual is directly responsible for the total project (i.e., no formal authority; committee solutions)* • *Does not provide the project-oriented emphasis necessary to accomplish the project tasks* • *Coordination becomes complex, and additional lead time is required for approval of decisions* • *Decisions normally favor the strongest functional groups* • *No customer focal point* • *Response to customer need is slow* • *Difficulty in pinpointing responsibilities* • *Motivation and innovation are decreased* • *Ideas tend to be functionally oriented with little regard for ongoing projects*

At the project level, the structure of a nonprojectized organization divides individually among all departments or functions within the project. The implementation features below are innovated or derived from literature review for the project level such as:

- Better control within a department or function of the project
- Variety of specialists can be hired
- Policies and responsibilities are clearly defined for each department
- Minimal risks and variations

- Communication channels are vertical or through upper-level management
- People require less project-management knowledge
- Lack of coordination and poor decision-making process

In summary, the implementation features of projectized or nonprojectized organizations were identified and could be applied at the project level. A successful implementation of organizational factors would improve the project performance [34]. It considers the authorities, responsibilities, and lines of control when organizing and integrating people into an effective work team. As a result, a well-organized work team would facilitate clear communication and management approaches.

3 Procurement Systems

Recent developments in procurement systems have focused on electronic-based procedures[14, 25, 26]. Theories of information systems have widely applied in procurement systems to simplify and enhance the process of procurement as well as legal aspects of a business [33]. Therefore, certain management strategies could be incorporated into procurement systems. This is a complementary and value added approach to the current developments in e-procurement systems.

Generally, a procurement system is a managerial structure that adopted by the client for the implementation, and at times eventual operation of a project [32]. The management of project procurement requires the contract management and change control to efficiently administer projects [40]. Project-delivery methods are related to the contract strategies used for the acquisition of goods or services involving the employer and the contractor.

Besides, lump-sum procurement system also describes as a fixed-price contract, which is the most common type of contract in the construction industry. This type of system always associates with the traditional approach to contracts, wherein competitive tendering and a bill of quantities commonly use in the project [42]. The contractor obliges to carry out the work at the negotiated contract value, and this type would provide the maximum cost protection for the employer if zero or minimal variations occur during the project [24]. The construction industry is full of risks and uncertainties. Procurement systems deal with risk allocation between the contractor and the employer. In a lump-sum contract, the risk allocation is regarded as fairer and more balanced in the perspective of employers because the employer has a better control in terms of the performance of the contractor and change management along the project. The roles and responsibilities are well defined and differentiated for the professionals who work in the project under this procurement system, particularly for the design–and-construction processes.

On the contrary, the design-and-build system overcomes the problem of having separate design and construction processes by incorporating them into a single organization [3]. This system requires the contractor to design and construct the project in a single contract package. It places the responsibility of saving the construction time and costs solely on the contractor. Usually, a design-and-build project is often associated with a negotiated approach of contract due to tight time requirements [20]. A competent contractor is selected through negotiations from among several short-listed contractors. Moreover, the design and construction can be carried out concurrently to shorten the project duration. One of primary reasons on why employers select the design–build procurement system is to shorten the overall project duration [11]. This ensures smooth cash flow and financial stability [23]. There are several pros and cons in relation to these two procurement systems. In general, the advantages and disadvantages of these methods could be summarized, as shown in Table 3 [42].

Table 3
Advantages and disadvantages of lump sum (traditional approach) and design and build systems [42]

Advantages	Disadvantages
Traditional approach:	
Competitive tendering is used	Decision process are slow and convoluted
Bill of quantities make for case of valuation of variations	Total project time is the longest of all options
High quality and functional standards are possible	It has low levels of buildability
There is cost certainty at the start of construction	Many organizational interfaces must be managed
Independent advice is given on most aspects of the process	
There are clear line of accountability	
A combination of best design and construction skills is possible	
There is flexibility for design changes	
Design and build	
There is single point of responsibility	There is very lack independent advice
Fixed price bids are used	Valuation of variations is not based on fixed price
Design and construction are integrated	It requires a detailed brief and client requirements at the outset
It has high levels of buildability	Changes can be expensive
Total project times are short	Client control of quality and functionality is minimized
Client involvement in the process can be minimized	Design and build firms lack of broad experience or expertise
Package deal systems offer off-the-shelf solutions	There are lower level of competition at tender than in the other approaches
There is competition on price and product	The tender process can be expensive to bidders
	Comparison of bids can be complicated
	It may not produce well-thought-out bids

Overall, procurement systems merely allocate risks and responsibilities between the prime contracting parties within a project. They focus on contract management. Therefore, it is important to look at other perspectives of managerial approach to enhance the effectiveness of procurement systems, particularly, the incorporation of suitable principles of projectized or nonprojectized organizations into procurement systems.

4 Development of A Mechanism for Analyzing the Compatibility of Different Procurement Systems and Organizational Forms

The McKinsey Seven S model was developed in the early 1980's by McKinsey and Company [37, 39]. This well-established framework explains and analyses both "hard" and "soft" aspects of organisational management such as strategy, structure, systems, shared values, style, staff, and skills. It is also closely related and applicable to a project delivery approach as the model is recognized for its completeness and comprehensiveness.

The shape of the model explains the interdependency of the variables as illustrated in Figure 1. The variables consider to be of critical importance to managers and practitioners [39].

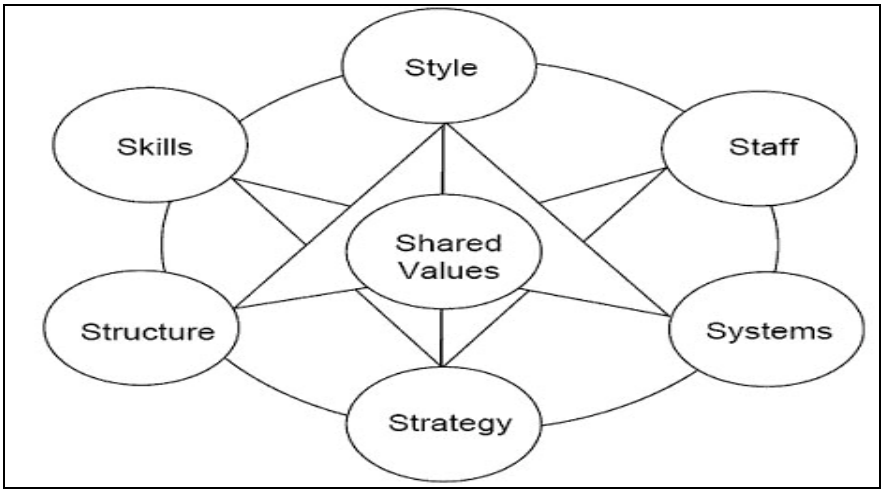


Figure 1
McKinsey Seven S Model [39]

Therefore, the Seven S "levers" was adopted in this review to organise the criteria of organizational management in order to assess the compatibility of the organisational forms (nonprojectized and projectized) with the selected procurement systems. They seem to be broad enough to encompass the critical aspects of implementing organisational management at the level of the project. In addition, they provide an important link to the business level of construction organisations where corporate and commercial strategic decisions are taken into consideration. As a result, it creates a significant impact on project level strategies and organisation. The details of the Seven S Model and its expansion and explanation for the construction project are as follow:

- **Strategy** describes as a plan or course of action in allocating resources to achieve identified goals over time [5, 45]. To the procurement systems as mentioned, they relate to the constraints for their design performance and requirements as well as overall project control.
- **Structure** defines as a skeleton of the organisation or the organisational chart [5, 45]. However, in construction project level, it is about the overall project delivery mechanism, particularly for organising the tasks involved and its layers of management.
- **Systems** are the routine processes and procedures followed within the organisation [30]. It is similar for the scope of works and job activities at the project level.
- **Shared values** refer to the significant meanings or guiding concepts that organisational members share [31, 39]. The culture of the overall management would determine the shared values (outcomes) generated from the other 6S at the project level.
- The way in which key managers behave in achieving organisational goals is considered to be the **style** variable; this variable is thought to encompass the cultural style of the organisation. In the construction project level, it is about the project leadership towards the authorities and communication among the workers or management teams.
- **Staff** describes as personnel categories within the organisation (e.g. engineers), whereas the skills variable refers to the capabilities of the staff within the organisation as a whole [41]. This relates to the level of capabilities required for the workforce at the project level.
- **Skills** refer to the actual level of skills and competencies of the employees working for the company. In the construction context, it is about the degree of specialization required for the project delivery approach.

Subsequently, the McKinsey 7 S model has been reviewed for checking the degree of compatibility for the strategic management organization and the procurement systems at the level of the construction project as shown in Table 4.

Table 4
Compatibility assessment for selected procurement systems and organisational forms utilising McKinsey Seven S Model

Degree of Compatibility **Low** **Neutral** **High** **1** **2** **3** **4** **5**			
Compatibility Criteria	**Procurement Systems**	**Organisational Form**	
Strategy		**Nonprojectized**	**Projectized**
Design performance and requirements	Lump Sum	5	1
	Design & Build	1	5
Overall project Control	Lump Sum	4	2
	Design & Build	1	5
Structure			
Complexity of tasks involved	Lump Sum	5	3
	Design & Build	3	5
Layers of management	Lump Sum	5	1
	Design & Build	1	5
Systems			
Scope of work	Lump Sum	5	2
	Design & Build	3	5
Procedures and standardization	Lump Sum	5	3
	Design & Build	1	5
Shared Values			
Type of culture – better project effectiveness and lower risks involved	Lump Sum	5	3

The comparison refers to general practice under a common scenario. According to the literature, the lump sum procurement system has a high compatibility level with the nonprojectized organization for all the 7S except for Skills. It is because a relatively high degree specialization is required for the procurement system compared to the nonprojectized organization in most cases, for example, a high rise and sophisticated building. Yet, the nonprojectized organization is quite relevant to the lump sum procurement system on less-complicated projects, which both promote a low degree of specialization for the routinized works.

On the other hand, the design and build system has a very high compatibility level with the projectized organization for all the 7 S. Both have a similar principle and approach in project delivery based on the review of the literature. However, the criteria of Structure, System, Staff and Skills were rated as neutral for the design and build with the nonprojectized organization. It is because these areas are subjected to further invesitgation on the actual project requirements and conditions. Overall, the cross comparison has distinguished a preliminary assessment of level of compatibility for the procurements systems and the organization principles.

5 Integration and Discussion

The existing procurement systems have served well within a project in terms of contractual arrangements and risk allocations [12, 27, 42]. However, procurement systems could be further improved, particularly at the perspective of its organizational form in order to have a more collaborative planning and organization at the outset of a project [19]. The organizational principles focus on the micro or project level. Different perspectives of organizational principles could be used to enhance the organizational approach to a project in the context of management.

In the context of construction management, project managers usually tend to select an appropriate procurement system for managing the project. It is very rare to have a well-organized form of organization for a particular contract arrangement in a project. Therefore, the integration herein aims to incorporate the organizational principles such that they work well in conjunction with procurement systems.

According to the review, the principles of projectized organization deem appropriate for design-and-build procurement systems. The features of a projectized organization are aligned with the design-and-build principles, where responsibilities of the contractor in terms of design and construction require a complete and unambiguous line of authority and communication structure within the project. When the project manager has a direct control over and response from the project participants, he or she could control or manage the project more effectively in the projectized organizational form. Moreover, these two methods also share the same shortcomings in terms of their principles, in which the design-and-build projects are costly and their technologies or constructions are typically sophisticated or completed on a one-time basis and are hardly flexible enough to be transferred to other projects.

Besides, the principles of nonprojectized organizations are well integrated with the lump-sum procurement system. The distinct departments and processes of the organizational form are able to streamline the project's effectiveness for the lump-sum contract when different specialists control their departments and perform their duties individually with a clear policy. This environment could produce a better product at a lower cost, where the risks or uncertainties are minimal and handled properly within each department. Nevertheless, the project requires long duration in this combination.

Apart from that, certain limitations of this empirical study need to be considered. The integration was carried out based on literature reviews by examining the fundamental characteristics of the projectized and nonprojectized organizational forms in a systematically approach, which highlighted each similarity and difference in terms of the compatibility for their principles. It has detailed the preliminary development and application of a mechanism for analyzing the

compatibility of construction procurement systems and organizational management forms. It has attempted to use the well established McKinsey 7 S model from general strategic management and applied it at the level of the construction project. Nevertheless, more works are required in developing the criteria under each of the interrelated 7 S to see whether the mechanism might be used as a generic tool for assessing compatibility.

The results would be more convincing if the integration could be tested and applied to real construction projects in the future. Meanwhile, It also could consider and promote integral management for the governance and management of enterprises [47, 48]. In addition, the study should extend to other organizational management strategies and forms and procurement systems, such as hybrid or matrix organization structures, cost-reimbursement systems, guaranteed maximum systems, and so on. These two limitations serve as the topics for future research.

Conclusion

Conventionally, procurement systems mreley focus on risk allocation and contract management. This paper highlights an innovative approach of incorporating the principles of projectized and nonprojectized organizations to selected procurement systems, namely, lump-sum and design-and–build systems. McKinsey 7 S model was adopted as a basic mechanism in reviewing the significant aspects of strategic management organization at the level of the construction project. It has identifed the key areas of compatibility between procurement systems and organizational forms.

The combined values were generated based on a comprehensive literature analysis. It provides a useful reference in construction project management. In summary, procurement systems can become more organized and effective by applying the organizational management strategies and structural forms of (a) design-and-build systems in conjunction with principles of projectized organizations, and (b) lump-sum systems in association with principles of nonprojectized organizations. The theory of incorporation renders an important insight for further enhancement of PMBOK from the viewpoints of both procurement systems and project organization theories.

References

[1] Alvarez, S. A., Paker, S. C. Emerging Firms and the Allocation of Control Rights: A Bayesian Approach. Academy of Management Review, 34, 209-227, 2009

[2] Arenius, M., Artto, K. A., Lahti, M., Meklin, J. Project Companies and the Multi-Project Paradigm: a New Management Approach. In: Slevine DP, Cleland DI, Pinto JK. (Eds.). The Frontiers of Project Management Research. Newtown Square: Project Management Institute, 2002

[3] Ashworth, A. Contractual Procedures in the Construction Industry. London: Pearson, 2006

[4] Benedek, P Compliance Management - A New Response to Legal and Business Challenges. Acta Polytechnica Hungarica, 9 (3), 135-148, 2012

[5] Boyle, S. Impact of Changes in Organisational Structure on Selected Key Performance Indicators for Cultural Organisations. International Journal of Cultural Policy, 13, 319-334, 2007

[6] Chan, A. P. C., Yung, E. H. K., Lam, P. T. I., Tam, C. M., Cheung, S. O. Application of Delphi Method in Selection of Procurement Systems for Construction Projects. Journal of Construction Management and Economics, 19, 699-718, 2001

[7] Chan, A. P. C. Evaluation of Enhanced Design and Build System: a Case Study of a Hospital Project. Journal of Construction Management and Economics, 18, 863-871, 2000

[8] Cheng, J. C. P., Law, K. H., Bjornsson, H., Jones, A., Sriram, R. A Service-oriented Framework for Construction Supply Chain Integration. Automation in Construction, 19, 245-260, 2010

[9] Cheung, S. O., Lam, T-I., Wan, Y-W. Lam, K-C. Improving Objectivity in Procurement Selection. Journal of Management in Engineering, 17, 132-139, 200

[10] Chiocchio, F., Beaulieu, G., Boudrias, J-S., Rousseau, V., Aube, C., Morin, E. M. The Project Involvement Index, Psychological Distress, and Psychological Well-Being: Comparing Workers from Projectized and Non-projectized Organizations. International Journal of Project Management, 28, 201-211, 2010

[11] Cho, K. M., Hyun, C. T., Koo, K. J., Hong, T. H. Partnering Process Model for Public-Sector Fast-Track Design-Build Projects in Korea. Journal of Management in Engineering, 26, 19-29, 2010

[12] Chong, H-Y., Zin, R. M. A Case Study into the Language Structure of Construction Standard form in Malaysia. International Journal of Project Management, 28:601-608, 2010

[13] Clifton, C., Duffield, C. F. Improved PFI/PPP Service Outcomes through the Integration of Alliance Principles. International Journal of Project Management, 24, 573-586, 2006

[14] Elbanna, A. From Intention to Use to Actual Rejection: The Journey of an e-procurement System. Journal of Enterprise Information Management, 23, 81-99, 2010

[15] Fortune, A, Mitchell, W. Unpacking Firm Exit at the Firm and Industry Levels: The Adaptation and Selection of Firm Capabilities. Strategic Management Journal, 33, 794-819, 2012

[16] Frigenti, E., & Comninos, D. The Practice of Project Management - A Guide to the Business-focused Approach. London: The Bath Press Ltd, 2002

[17] Gluch, P. Unfolding Roles and Identities of Professionals in Construction Projects: Exploring the Informality of Practices. Journal of Construction Management and Economics, 27, 959-968, 2009

[18] Gulati, R, Puranam, P., Tushman, M. Meta-Organization Design: Rethinking Design in Interorganizational and Community Contexts. Strategic Management Journal, 33, 571-586, 2012

[19] Gunter, H., Grote, G. Collaborative Planning and its Antecedents: An Assessment in Supply Chain Relationships. Journal of Management and Organization, 18, 36-52, 2012

[20] Halpin, D. W. Construction Management. New York: John Wiley & Sons, 2006

[21] Ive, G., Chang, C-Y. The Principle of Inconsistent Trinity in the Selection of Procurement Systems. Journal of Construction Management and Economics, 25, 677-690, 2007

[22] Jost, P-J., Lammers, F. The organization of Project Evaluation under Competition. Review of Managerial Science, 3, 141-155, 2009

[23] Kaplanoglu, S. B., Arditi, D. Guidelines for Pre-Project Peer Reviews in Construction Contracting. International Journal of Project Organisation and Management, 2, 154-173, 2010

[24] Kerzner, H. Project Management: A Systems Approach to Planning, Scheduling and Controlling. New Jersey: John Willey & Sons, 2006

[25] Ketikidis, P. H., Kontogeorgis, A., Stalidis, G., Kaggelides, K. Applying e-procurement System in the Healthcare: The EPOS paradigm. International Journal of Systems Science, 41, 281-299, 2010

[26] Liu, Q., Sun, S. X., Wang, H., Zhao, J. A Multi-Agent-based System for e-procurement Exception Management. Knowledge-Based Systems, 24, 49-57, 2011

[27] Lumineau, F., Quélin, B. V. An Empirical Investigation of Interorganizational Opportunism and Contracting Mechanisms. Strategic Organization, 10, 55-84, 2012

[28] Luu, D. T., Ng, S. T., Chen, S. E., Jefferies, M. A Strategy for Evaluating a Fuzzy Case-based Construction Procurement Selection System. Advances in Engineering Software, 37, 159-171, 2006

[29] Luu, D. T., Ng, S. T., Chen, S. E. Parameters Governing the Selection of Procurement System - An Empirical Survey. Engineering, Construction and Architectural Management,10, 209-218, 2003

[30] Lynch, R. Corporate Strategy, 4th edition. Prentice Hall: UK, 2005

[31] Martins, E., Terblanche, F. Building Organisational Culture that Stimulates Creativity and Innovation. European Journal of Innovation Management, 6, 64-74, 2003

[32] Masterman, J. W. E. An Introduction to Building Procurement Systems. Spon Press: New York, 2002

[33] Mishra, A., Mishra, D A Legal Business Information System: Implementation Process Context. Acta Polytechnica Hungarica , 8 (2), 45-59, 2011

[34] Mollick, E. People and Process, Suits and Innovators: The Role of Individuals in Firm Performance. Strategic Management Journal, 33, 1001-1015, 2012

[35] Mueller, J. The Interactive Relationship of Corporate Culture and Knowledge Management: A Review. Review of Managerial Science, 6, 183-201, 2012

[36] Oyegoke, A. S., Kiiras, J. Development and Application of the Specialist Task Organization Procurement Approach. Journal of Management in Engineering, 25, 131-142, 2009

[37] Pascale, R., Athos, A. The Art of Japanese Management. Penguin Books: London, 1981

[38] Patrick, F. Multi-Project Constraint Management: the ‘‘Critical Chain” Approach. In: Dinsmore PC, Cabanis-Brewin J. (Eds.), The AMA Handbook of Project Management. Amacom American ManagementL: New York, 2006

[39] Peters, T., Waterman, R. In Search of Excellence. Harper & Row: New York, 1982

[40] PMI (2008) A Guide to the Project Management Body of Knowledge, Project Management Institute: Pennsylvania, 2008

[41] Purcell, J., Boxal, P. Strategy and Human Resource Management (Management, Work and Organisations. Palgrave Macmillan: UK, 2003

[42] Rowlinson, S. M., McDermott, P. Procurement Systems: A Guide to Best Practice in Construction. E & FN Spon: London, 1999

[43] Scherpereel, C. M. The Option-Creating Institution: A Real Options Perspective on Economic Organization. Strategic Management Journal, 29, 455-470, 2008

[44] Turner, J. R. The Handbook of Project Based Management. MrGraw-Hill: New York, 2009

[45] Waterman, R. Jr., Peters, T., Phillips, JR. Structure Is Not Organisation. Business Horizons, 23,14-26, 1980

[46] Yoo, J. W., Kim, K. Board Competence and the Top Management Team's External Ties for Performance. Journal of Management and Organization, 18, 142-158, 2012

[47] Belak, J., Duh, M. Integral Management: Key Success Factors in the MER Model. Acta Polytechnica Hungarica, 9, 5-26, 2012

[48] Belak, J., Milfelner, B. Enterprise Culture as One of the Enterprise's Key Success Factors (Integral Management Approach): Does the Internal and External Cultural Orientation Matter? Acta Polytechnica Hungarica, 9,27-44, 2012

Determination of the Necessary Number of Technicians on the Faculty

Tomislav Šuh[1], Dragan Mitić[2], Dragica Lebl-Antonić[3], Aleksandar Lebl[4]

[1] Institute for Telecommunications and Electronics, IRITEL A.D. BELGRADE Batajnički put 23, 11080 Belgrade, Serbia; E-mail: suh@iritel.com

[2] Institute for Telecommunications and Electronics, IRITEL A.D. BELGRADE Batajnički put 23, 11080 Belgrade, Serbia; E-mail: mita@iritel.com

[3] Faculty of Pharmacy, Vojvode Stepe 450, 11000 Belgrade, Serbia; E-mail: dragica@pharmacy.bg.ac.rs

[4] Institute for Telecommunications and Electronics, IRITEL A.D. BELGRADE Batajnički put 23, 11080 Belgrade, Serbia; E-mail:lebl@iritel.com

Abstract: In this paper we present the mathematical analysis of engagement of technicians in the teaching (educational) process on the faculty. The technicians receive the requests from professors and assistants. In the case that all technicians are busy in the moment of generating a new request, some of the assistants take over processing of the request. We analyse the probability of unwanted states, i.e. the states when it was necessary to engage the assistants on processing the request, which could have been processed by the technicians, if they had been free. The number of engaged technicians is determined in such a way that the probability of these unwanted states is satisfactorily small, i.e. that these states take available assistants' time in reasonable limits.

Keywords: queueing systems; teaching process; unwanted states

1 Introduction

Modern technology progress affects all aspects of human life. Among others, this progress is also visible in the faculty teaching (educational) process, where the efforts are made to implement modern methods of students' learning and testing [1]. Besides this, the great attention is devoted to the quality of faculty staff [2] and, indirectly, to the teaching quality. The quality of teaching process is not based only on the quality of teaching staff (professors and assistants), but also on the quality of the technicians support.

Practical preparation and realization of the teaching process on many faculties is connected with engagement of technicians, who achieve contacts with students and prepare the necessary material for the teaching process. Technicians receive requests from professors, but also from assistants. Their engagement represents random process from the aspect of the requests, which they receive from professors and assistants, and from the aspect of necessary engagement time for each request. There are situations when all available technicians are engaged and the new request, which is generated in that moment, could not be processed (serviced). In such a situation it is possible that some assistant assumes execution of the request, which would have been done by the technician, if some technician had been free. This situation is not desirable, because it occupies the available time of the assistant, instead of using this time for assistant's main activity. But, if it doesn't happen too often, it can be reasonable and may be allowed to complete the necessary requests in time, instead of engaging new technicians. The analysis of the need for employing new technicians depends on the acceptable probability level that the assistants are engaged on processing (servicing) the requests.

In this paper we present mathematical model of teaching realization on the faculty. In the practice professors and assistants generate requests, which technicians process, and in the case that all technicians are busy, assistants take over the execution of these requests.

2 Theoretical Base

Queueing systems are often analyzed in literature [3], [4]. There are many contributions considering these systems. One interesting survey of different queueing systems is presented in [5]. The procedure of system defining in one specific case is presented in [6].

Queueing systems are described and analyzed using system states, which can be defined by the number of requests in the system, number of users, who generate requests in system, or the number of processing channels. The system, defined in one of these three ways, can be, for example, in some state i, which is designated as $\{i\}$.

Let us consider the part of one system, presented in Figure 1, which consists of three states. In some moment t the system is in state $\{i\text{-}1\}$, or in state $\{i\}$, or in state $\{i+1\}$. Let us consider only the events connected with the state $\{i\}$. During the time Δt (i.e. during the time interval $t - (t+\Delta t)$) the following events are possible:

- system is in moment t in state $\{i\text{-}1\}$; during the time interval Δt system passes into state $\{i\}$; the probability of this event can be expressed as $\lambda_{i-1} \cdot \Delta t$, where λ_{i-1} is the arrival rate of new requests (intensity of new requests generation) in state $\{i\text{-}1\}$;

- system is in moment t in state $\{i\}$; during the time interval Δt system passes into state $\{i+1\}$; the probability of this event can be expressed as $\lambda_i \cdot \Delta t$;
- in the moment t system is in state $\{i\}$ and during the time interval Δt passes into the state $\{i-1\}$; the probability of this event is $\mu_i \cdot \Delta t$, where μ_i is the service rate (intensity of processing the requests) in state $\{i\}$;
- in the moment t system was in state $\{i\}$, and after time interval Δt remains in the same state; the probability of this event is $1-\lambda \cdot \Delta t-\mu_i \cdot \Delta t$;
- in the moment t system is in state $\{i+1\}$ and during the time interval Δt passes into the state $\{i\}$; the probability of this event is $\mu_{i+1} \cdot \Delta t$.

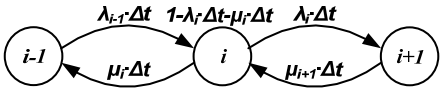


Figure 1
Diagram of possible states and possible transitions in the system

Let $P_{i-1}(t)$, $P_i(t)$ and $P_{i+1}(t)$ present the probability that the system is in state $\{i-1\}$, $\{i\}$ and $\{i+1\}$, respectively. The probability that the system is in moment $t+\Delta t$ in the state $\{i\}$ can be expressed as:

$$P_i(t+\Delta t) = P_i(t) \cdot (1-\lambda_i \cdot \Delta t - \mu_i \cdot \Delta t) + P_{i-1}(t) \cdot \lambda_{i-1} \cdot \Delta t + P_{i+1}(t) \cdot \mu_{i+1} \cdot \Delta t \tag{1}$$

In the limiting case, when $\Delta t \to 0$, we obtain the differential equation:

$$\frac{dP_i(t)}{dt} = -P_i(t) \cdot (\lambda_i + \mu_i) + P_{i-1}(t) \cdot \lambda_{i-1} + P_{i+1}(t) \cdot \mu_{i+1} \tag{2}$$

The equations of this kind can be written for all possible states $\{i\}$ of the system, or in a matrix form:

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{P}(t) \cdot \mathbf{A} \tag{3}$$

where $\mathbf{P}(t)$ is the vector of state probabilities, and $\mathbf{A}$ is the matrix of transition intensities between the system states. In matrix $\mathbf{A}$ the elements on the diagonal represent the intensities of „outgoing“ from the corresponding state, and are taken with the sign „-“. The other elements in the matrix represent the intensities of „incoming“ to the corresponding state, and are taken with the sign „+“.

After the sufficiently long period of time (i.e. when $t \to \infty$), the probabilities of system states tend to stationary values, i.e.:

$$\frac{d\mathbf{P}(t)}{dt} = 0 \tag{4}$$

According to (3) and (4), the system of differential equations becomes the system of linear equations:

$$\mathbf{0} = \mathbf{P} \cdot \mathbf{A} \tag{5}$$

The additional equation, which is necessary to solve this system of linear equations, is:

$$\sum P_i = 1 \tag{6}$$

where $\sum P_i$ includes probabilities of all possible states of the analyzed system.

3 Analytical Model

Let us analyse, in general case, one system, presented in Figure 2. The main parameters of this system are:

- k – number of professors, who generate requests in the system;
- m – number of assistants, who generate requests and, if necessary, process generated request;
- n – number of technicians, who process generated requests;
- λ – mean arrival rate expressed as number of arrivals in one second (s^{-1}), i.e. mean intensity of requests generation by each professor and each assistant;
- μ_a – mean assistants' service rate in s^{-1}, i.e. mean intensity of requests processing by each assistant;
- μ_t – mean technicians' service rate in s^{-1}, i.e. mean intensity of requests processing by each technician.

In this system $\{i,j\}$ presents the state where requests are processed by i assistants and j technicians (for example, the state $\{1,2\}$ is the state where requests are processed by 1 assistant and 2 technicians).

Let us now analyse only one state $\{i,j\}$ (again $\{1,2\}$), Figure 2. In this state the new request may be generated by k professors and m-1 assistants. When this new request is generated (with total intensity $(k+m-1)\cdot\lambda$), the system passes to the state $\{1,3\}$.

We can come to state $\{1,2\}$ from the state $\{1,1\}$ in such a way that, also, k professors and m-1 assistants generate new request with the intensity $(k+m-1)\cdot\lambda$.

The state $\{1,2\}$ is left when the processing of some request is finished. If an assistant finishes the processing, the system passes to the state $\{0,2\}$ (the intensity

of this event is μ_a). If one of two active technicians finishes the processing, the system passes to the state {1,1} (the intensity of this event is $2\cdot\mu_t$).

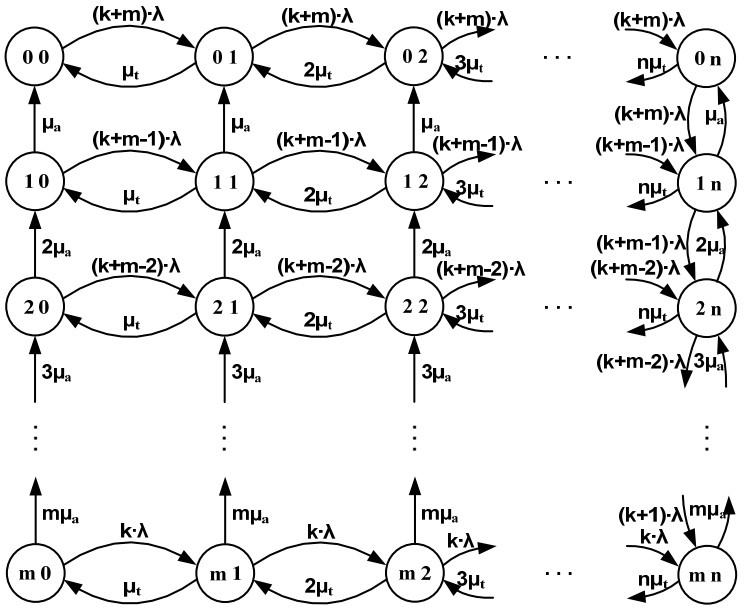


Figure 2

The model of teaching process on the faculty with *k* professors, *m* assistants and *n* technicians

We can come to the state {1,2} from the state {1,3} in such a way that one of three active technicians finishes processing of the request (this happens with the intensity $3\cdot\mu_t$), or from the state {2,2} in such a way that one of two assistants, who processes the request, finishes the processing of his request (this happens with the intensity $2\cdot\mu_a$).

The desirable states in the system are the ones, when assistants are not engaged on the processing of the requests, but the technicians succeed to process all requests, and the assistants have time for more important jobs. The result of these jobs is, among others, the generation of new requests. The undesirable states are the ones, when the assistants are engaged on the processing of the requests. The especially undesirable states are the ones, when the assistants are engaged on processing the requests, although one or more technicians are not engaged. The system can come to these states when we do not use the principle that free technicians take over the processing of the requests, which assistants started to process.

Let p_{ij} presents the probability of state $\{i,j\}$, i.e. the probability of the state in which *i* assistants and *j* technicians are processing the requests. The probability of undesirable states (i.e. states when assistants are engaged on the processing of requests) can be calculated according to the formula:

$$p_u = 1 - \sum_{j=0}^{n} p_{0j} \tag{7}$$

The probability of especially undesirable states (i.e. states when assistants are engaged on the processing of requests, although there are free technicians) can be calculated using the formula:

$$p_{uft} = 1 - \sum_{j=0}^{n} p_{0j} - \sum_{i=1}^{m} p_{in} \tag{8}$$

Table 1 presents the matrix of the transition intensities between the states of the system from Figure 2. When considering one transition, the initial state is read in the first column of the Table 1, and the final state is read in the first row of the Table 1.

4 Results of the Analysis

On the basis of the theoretical presentation from previous section, we shall analyse the system with the following characteristics:

- there are two professors in the system and they generate the requests;
- there is one technician in the system, who is engaged on processing the requests;
- there are 4 assistants in the system, and they can generate requests, and, if necessary, they can also process the requests;
- technician processes the requests with the mean intensity μ_t;
- assistants process the requests with the mean intensity μ_a;
- professors and assistants generate requests with the mean intensity λ;
- if some assistant is engaged on processing the request, he can not generate the requests;
- $\{i,j\}$ is the state of the system when there are i technicians and j assistants, who are processing the requests;
- technician is the first to process the requests;
- if the technician is already occupied processing the request, assistants take over processing the new request.

Table 1
Matrix of the transition intensities between the states of the system with *n* technicians

	00	01	...	0n	10	11	...	1n	20	21	...	2n	...	m0	m1	...	mn
00	$-(k+m)\cdot\lambda$	μ_t	...	0	μ_a	0	...	0	0	0	...	0	...	0	0	...	0
01	$(k+m)\cdot\lambda$	$-(k+m)\cdot\lambda-\mu_t$	...	0	0	0	...	0	0	0	...	0	...	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0n	0	0	...	$-(k+m)\cdot\lambda-n\cdot\mu_t$	0	0	...	$(k+m)\cdot\lambda$	0	0	...	0	...	0	0	...	0
10	0	0	...	0	$-(k+m-1)\cdot\lambda-\mu_a$	$(k+m-1)\cdot\lambda$	...	0	0	0	...	0	...	0	0	...	0
11	0	0	...	0	μ_t	$-(k+m-1)\cdot\lambda-\mu_t-\mu_a$	...	0	0	0	...	0	...	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1n	0	0	...	$(k+m)\cdot\lambda$	0	0	...	$-(k+m-1)\cdot\lambda-\mu_a-n\cdot\mu_t$	0	0	...	$(k+m-1)\cdot\lambda$	...	0	0	...	0
20	0	0	...	0	0	0	...	0	$-(k+m-2)\cdot\lambda-2\cdot\mu_a$	$(k+m-2)\cdot\lambda$	...	0	...	0	0	...	0
21	0	0	...	0	0	0	...	0	μ_t	$-(k+m-2)\cdot\lambda-2\cdot\mu_a-\mu_t$	...	0	...	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2n	0	0	...	0	0	0	...	$(k+m-1)\cdot\lambda$	0	0	...	$-(k+m-2)\cdot\lambda-2\cdot\mu_a-n\cdot\mu_t$	...	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
m0	0	0	...	0	0	0	...	0	0	0	...	0	...	0	μ_t	...	0
m1	0	0	...	0	0	0	...	0	0	0	...	0	...	$-k\cdot\lambda-m\cdot\mu_a$	$-k\cdot\lambda-\mu_t-m\cdot\mu_a$	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
mn	0	0	...	0	0	0	...	0	0	0	...	0	...	0	0	...	$-n\cdot\mu_t-m\cdot\mu_a$

Figure 3 presents the model, which illustrates the example where one technician processes the requests, and the matrix of the transition probabilities in that case is presented in Table 2.

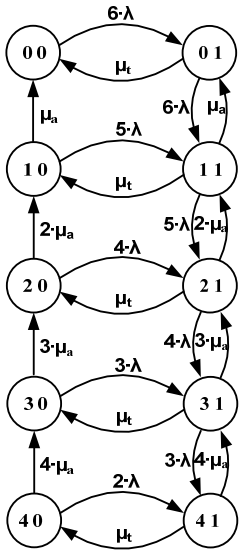


Figure 3
The model of the analyzed system with two professors, four assistants and one technician

Table 2
Matrix of the transition intensities between the states of the system with one technician

	00	01	10	11	20	21	30	31	40	41
00	$-6\cdot\lambda$	μ_t	μ_a	0	0	0	0	0	0	0
01	$6\cdot\lambda$	$-6\cdot\lambda-\mu_t$	0	μ_a	0	0	0	0	0	0
10	0	0	$-5\cdot\lambda-\mu_a$	μ_t	$2\cdot\mu_a$	0	0	0	0	0
11	0	$6\cdot\lambda$	$5\cdot\lambda$	$-5\cdot\lambda-\mu_t-\mu_a$	0	$2\cdot\mu_a$	0	0	0	0
20	0	0	0	0	$-4\cdot\lambda-2\cdot\mu_a$	μ_t	$3\cdot\mu_a$	0	0	0
21	0	0	0	$5\cdot\lambda$	$4\cdot\lambda$	$-4\cdot\lambda-\mu_t-2\cdot\mu_a$	0	$3\cdot\mu_a$	0	0
30	0	0	0	0	0	0	$-3\cdot\lambda-3\cdot\mu_a$	μ_t	$4\cdot\mu_a$	0
31	0	0	0	0	0	$4\cdot\lambda$	$3\cdot\lambda$	$-3\cdot\lambda-3\cdot\mu_a-\mu_t$	0	$4\cdot\mu_a$
40	0	0	0	0	0	0	0	0	$-2\cdot\lambda-4\cdot\mu_a$	μ_t
41	0	0	0	0	0	0	0	$3\cdot\lambda$	$2\cdot\lambda$	$-4\cdot\mu_a-\mu_t$

According to Figure 3, the desirable states in this system are {0,0} and {0,1}. All other states are undesirable, and the probability of this situation (p_u) can be calculated using the formula (7).

Figure 4 presents the probabilities p_u for the analyzed system with one technician. The probabilities of these states are read on y axis, and the intensity μ_t is read on the x axis. The parameter for the presented curves is intensity μ_a (μ_a ranges from 1 to 5). The curves are presented for two values of intensities of requests generation: λ=0.1 and λ=0.5.

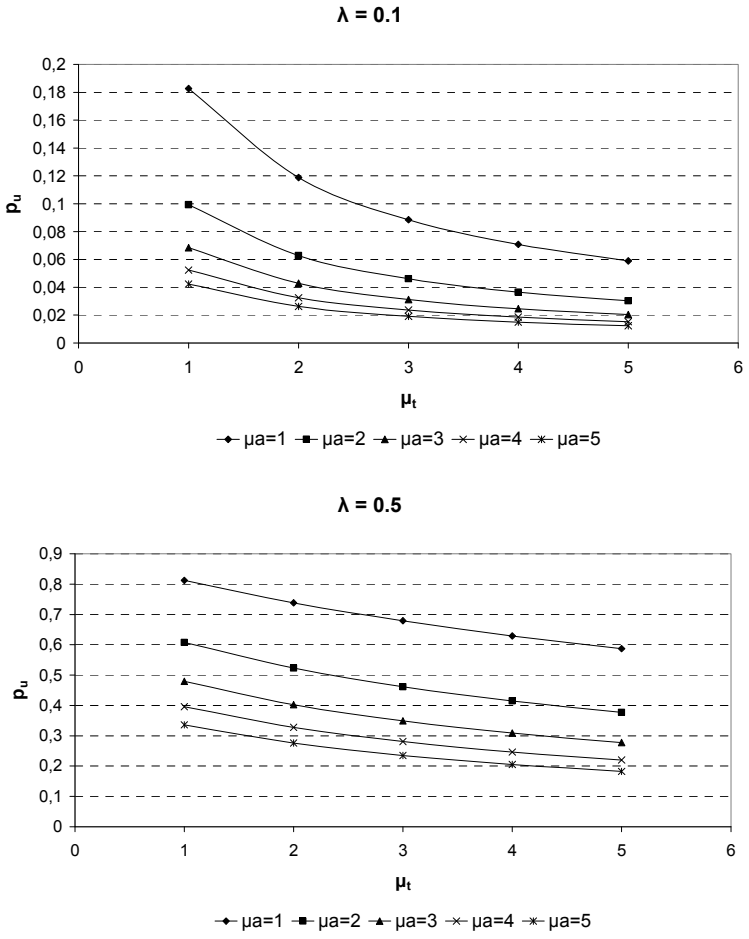


Figure 4
Probabilities of undesirable states in the system with one technician, in which assistants must be engaged on processing the requests

In the system, which is here analyzed, the states {1,0}, {2,0}, {3,0} and {4,0} are especially undesirable. The probability of such a state (p_{ufi}), in which the assistants are engaged on the processing of the requests, and the technician is free, can be calculated using the formula (8).

Figure 5 presents the probability of especially undesirable state p_{ufi} in the system with one technician.

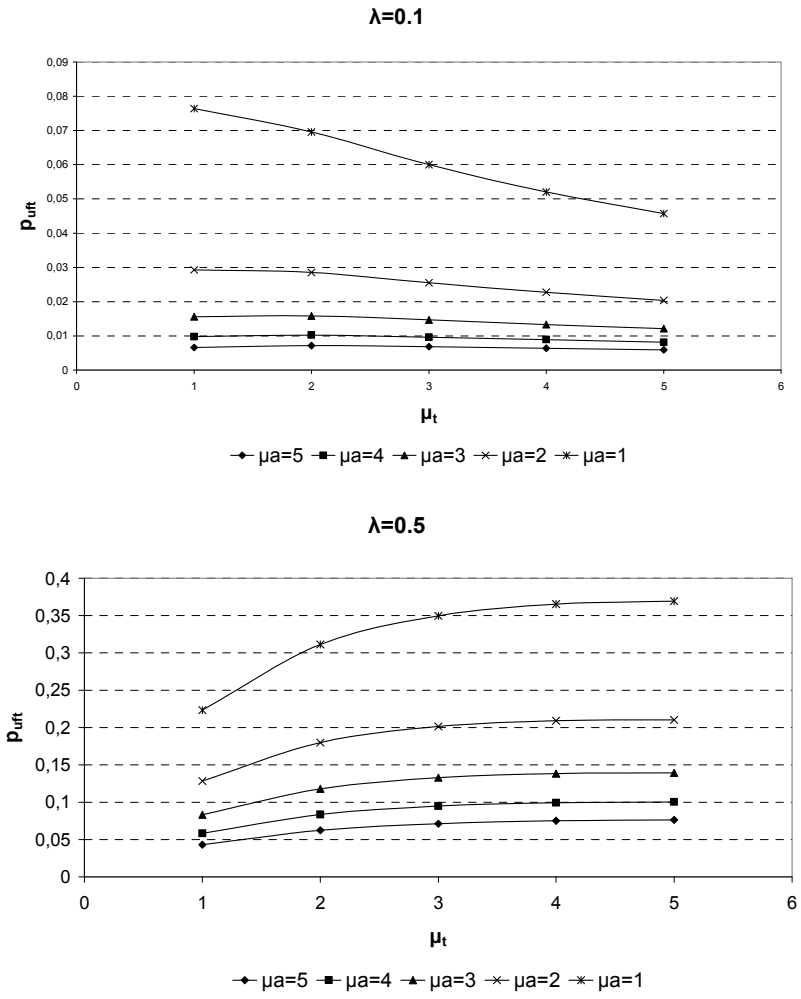


Figure 5
Probabilities of especially undesirable states in the system with one technician

We shall compare the presented results for the system with one technician to the results for the system of the same type, with the only difference that two technicians are engaged on the processing of the requests.

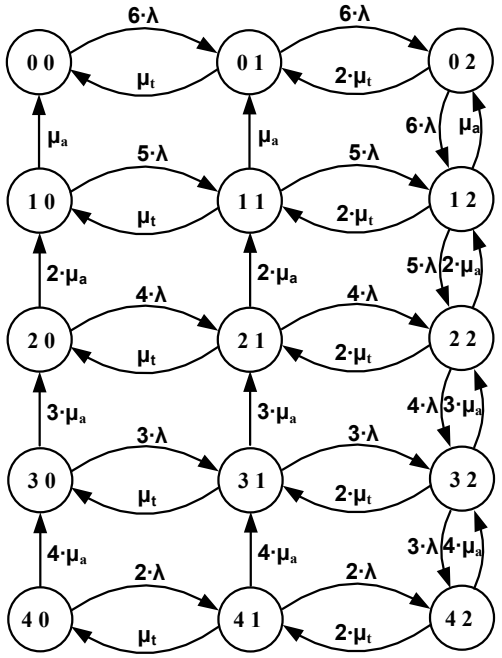


Figure 6

The model of the analyzed system with two professors, four assistants and two technicians

Figure 6 presents the model, which illustrates the analyzed example, where two technicians process the requests, and the matrix of transition probabilities in that case is presented in Table 3.

According to Figure 6, in the system, which is analyzed, the desirable states are {0,0}, {0,1} and {0,2}. All other states are undesirable, and the probability of such a situation (p_u) is calculated, again, using the formula (7).

Figure 7 presents the probabilities p_u for the analyzed system with two technicians.

The probability of especially undesirable state (p_{uft}), in which the assistants are engaged on the processing of requests although one or both technicians are free, is calculated using formula (8).

Figure 8 presents the probability p_{uft} in the system with two technicians.

Table 3

Matrix of the transition intensities between the states of the system with two technicians

	00	01	02	10	11	12	20	21	22	30	31	32	40	41	42
00	$-6\cdot\lambda$	μ_t	0	μ_a	0	0	0	0	0	0	0	0	0	0	0
01	$6\cdot\lambda$	$-6\cdot\lambda-\mu_t$	$2\cdot\mu_t$	0	μ_a	0	0	0	0	0	0	0	0	0	0
02	0	$6\cdot\lambda$	$-6\cdot\lambda-2\cdot\mu_t$	0	0	μ_a	0	0	0	0	0	0	0	0	0
10	0	0	0	$-5\cdot\lambda-\mu_a$	μ_t	0	$2\cdot\mu_a$	0	0	0	0	0	0	0	0
11	0	0	0	$5\cdot\lambda$	$-5\cdot\lambda-\mu_t-\mu_a$	$2\cdot\mu_t$	0	$2\cdot\mu_a$	0	0	0	0	0	0	0
12	0	0	$6\cdot\lambda$	0	$5\cdot\lambda$	$-5\cdot\lambda-2\cdot\mu_t-\mu_a$	0	0	$2\cdot\mu_a$	0	0	0	0	0	0
20	0	0	0	0	0	0	$-4\cdot\lambda-2\cdot\mu_a$	μ_t	0	$3\cdot\mu_a$	0	0	0	0	0
21	0	0	0	0	0	$5\cdot\lambda$	$4\cdot\lambda$	$-4\cdot\lambda-\mu_t-2\cdot\mu_a$	$2\cdot\mu_t$	0	$3\cdot\mu_a$	0	0	0	0
22	0	0	0	0	0	0	0	$4\cdot\lambda$	$-4\cdot\lambda-2\cdot\mu_t-2\cdot\mu_a$	0	0	$3\cdot\mu_a$	0	0	0
30	0	0	0	0	0	0	0	0	0	$-3\cdot\lambda-3\cdot\mu_a$	μ_t	0	$4\cdot\mu_a$	0	0
31	0	0	0	0	0	0	0	0	$3\cdot\lambda$	$-3\cdot\lambda-3\cdot\mu_a-\mu_t$	$2\cdot\mu_t$	0	0	$4\cdot\mu_a$	0
32	0	0	0	0	0	0	0	0	$4\cdot\lambda$	$3\cdot\lambda$	$-3\cdot\lambda-2\cdot\mu_t-3\cdot\mu_a$	0	0	0	$4\cdot\mu_a$
40	0	0	0	0	0	0	0	0	0	0	0	$-2\cdot\lambda-4\cdot\mu_a$	μ_t	$-2\cdot\lambda-4\cdot\mu_a-\mu_t$	$2\cdot\mu_t$
41	0	0	0	0	0	0	0	0	0	0	$4\cdot\mu_a$	0	$2\cdot\lambda$	$-2\cdot\lambda-4\cdot\mu_a-\mu_t$	$2\cdot\mu_t$
42	0	0	0	0	0	0	0	0	0	0	0	$4\cdot\mu_a$	0	$2\cdot\lambda$	$-2\cdot\mu_t-4\cdot\mu_a$

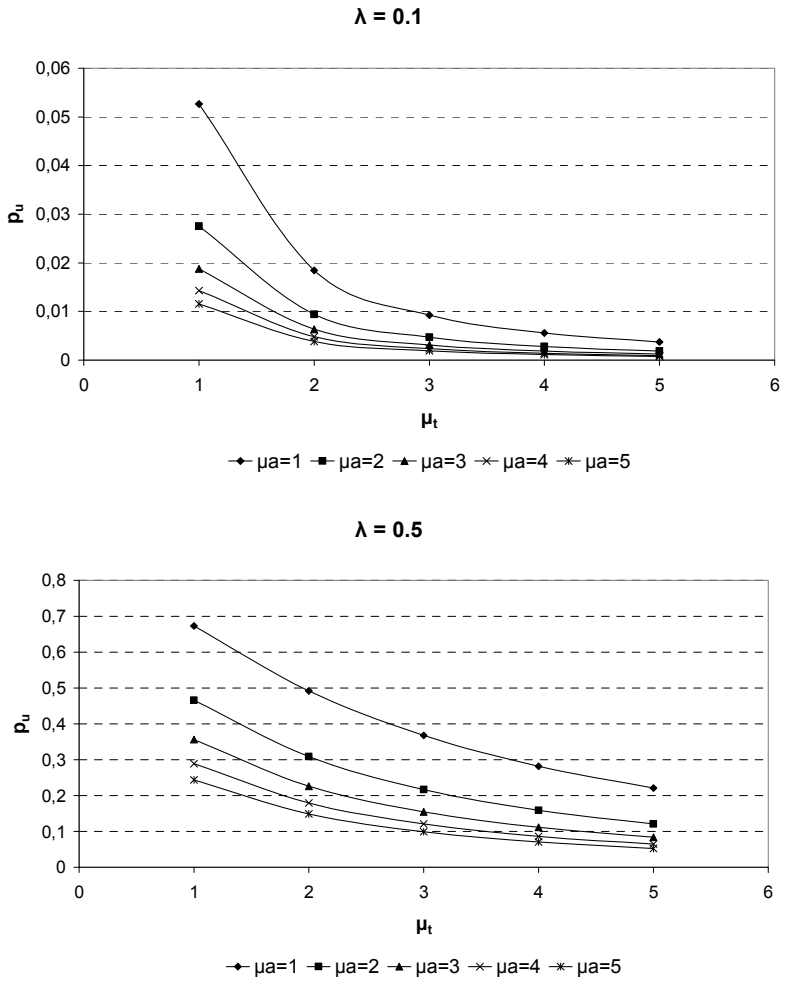


Figure 7
The probability of undesirable states in the system with two technicians, in which assistants must be engaged on processing the requests

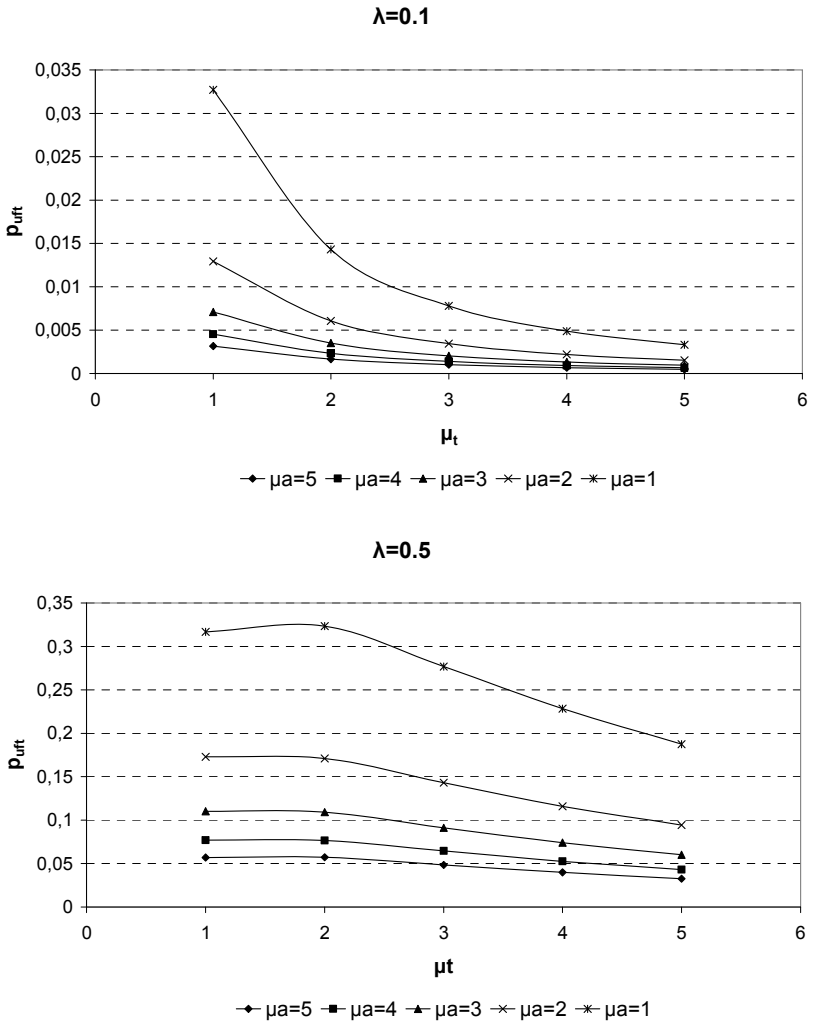


Figure 8
The probabilities of especially undesirable states in the system with two technicians

5 Analysis of the Calculated Results

Comparing the probabilities of assistants' engagement in the cases when there is one and when there are two technicians in the system, we can say that it is possible to reduce the probability of assistants' engagement more than twice when there are two technicians in the system instead of one technician, although the intensity of processing the requests is the same in both cases. For example, if λ=0.1, μ_t=1, μ_a=1, the probability of assistants' engagement is 0.18 when there is one technician in the system (Figure 4), and 0.052 when there are two technicians in the system (Figure 7). This is much better than in the case when the technician processes the requests two times faster (μ_t=2), and the assistants continue to process the requests with the same intensity (μ_a=1), because in that case the probability of assistants' engagement on processing requests is 0.12 (Figure 4).

Similar results are obtained comparing the probabilities of especially undesirable states, in which assistants are engaged on the processing of requests, although there are free technicians. For example, if λ=0.1, μ_t=1, μ_a=1 the probability of these, especially undesirable states is 0.077 when there is one technician in the system (Figure 5), and 0.033 when there are two technicians in the system (Figure 8). This is much better than in the case when the technician processes the requests two times faster (μ_t=2), and the assistants continue to process the requests with the same intensity (μ_a=1), because in that case the probability of especially undesirable states is 0.07 (Figure 5).

If it is λ=0.5 and μ_t<3 in the case of one technician, or μ_t<1.5 in the case of two technicians, system functions in the state where the traffic of generated requests is greater than the traffic, which can be processed by the technicians, i.e.:

$$\frac{(k+m)\cdot\lambda}{n\cdot\mu_t} > 1 \tag{9}$$

The state, when the condition expressed by formula (9) is satisfied, is known in the theory as the state when the system is unstable.

The graphs of especially undesirable states in the case that λ=0.5 have, seemingly, illogical shape, because the graphs are growing functions (i.e., when the technicians process the request faster, the probability of especially undesirable states increases). If the system functions in these conditions, it is not dimensioned well, i.e. it is necessary to avoid system operation in the state where total intensity of requests generation is greater than total intensity of technicians' processing of the requests.

Conclusion

In this paper we present the method for determination of required number of technicians for the teaching process implementation on the faculty. The technicians are engaged on the processing of the professors' and assistants'

requests. If in the moment of generating the new request there are no free technicians, processing of the request takes over one of the assistants. It is important that such an event is as rare as possible, because in that case assistant can't do his regular job. On the other hand, if there are too many technicians, they are not adequately engaged on their job, i.e. they will have free time. The required number of technicians is situated between these two extreme cases. This required number of technicians can be determined depending on the allowed level of assistants' engagement for processing the requests (which also could have been done by the technicians, if they had been free).

References

[1] Maravić Čisar, S., Radosav, D., Markoski, B., Pinter, R., Čisar, P.: Computer Adaptive Testing of Student Knowledge, in Acta Polytechnica Hungarica, Vol. 7, No. 4, 2010, pp. 139-152

[2] Stoklasa, J., Talašová, J., Hoček, P.: Academic Staff Performance Evaluation – Variants of Models, in Acta Polytechnica Hungarica, Vol. 8, No. 3, 2011, pp. 91-111

[3] Kleinrock, L.: Queueing Systems, John Wiley & Sons, New York, 1975

[4] Teletraffic Engineering Handbook: ITU-D SG 2/16&ITC, Draft 2001-06-20

[*5*] Filipowitz, B., Kwiecien, J.: Queueing Systems and Networks. Models and Applications, Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol. 56, No. 4, December 2008, pp. 379-390

[6] Petrović, G., Petrović, N., Marinković, Z.: Application of the Markov Theory to Queueing Networks, Facta Universitatis, Series Mechanical Engineering, Vol. 6, No. 1, 2008, pp. 45-56

ANFIS-based Indoor Location Awareness System for the Position Monitoring of Patients

Chih-Min Lin[1], Yi-Jen Mon[2], Ching-Hung Lee[3], Jih-Gau Juang[4], Imre J. Rudas[5]

[1]Department of Electrical Engineering, Yuan Ze University, Chung-Li, Taoyuan, 320, Taiwan, e-mail: cml@saturn.yzu.edu.tw

[2]Department of Computer Science and Information Engineering, Taoyuan Innovation Institute of Technology, Chung-Li, Taoyuan, 320, Taiwan e-mail: monbuy@tiit.edu.tw

[3]Department of Mechanical Engineering, National Chung Hsing University, Taichung, 402, Taiwan, e-mail: chleenchu@dragon.nchu. edu.tw

[4]Department of Communications, Navigation and Control Engineering, National Taiwan Ocean University, Keelung, 202, Taiwan, e-mail: jgjuang@ntou.edu.tw

[5]Óbuda University, Bécsi út 96/B, H-1034 Budapest, Hungary, e-mail: rudas@uni-obuda.hu

Abstract: In this paper, an adaptive network based fuzzy inference system (ANFIS) combining a location awareness system (LAS) is used to to develop an application of patient's position monitoring system. The setup procedure of LAS is carried out in advanced, then the ANFIS is designed to demonstrate the performance of the proposed methodology. From experimental and simulation results, satisfactory performance of LAS for patient's position monitoring can be achieved.

Keywords: Wireless sensor network (WSN); Location awareness system (LAS); ZigBee; ANFIS

1 Introduction

The purpose of this paper is to construct a location awareness system (LAS) [1] based on an adaptive network based fuzzy inference system (ANFIS) [2]. The technology of LAS is based on the protocol of wireless sensor network (WSN) [3-7]. The LAS is composed of the gateway, location nodes, server, client monitor, controller unit, etc. A variety of control techniques can be achieved by using the advantages of low cost and low power consumption technology of WSN. Many

applications of WSN have been developed successfully. The well known WSN is based on the IEEE 802.15.4/ZigBee standard to define the layers of network so as to achieve benefits of low cost and low power consumption.

The indoor LAS uses the received signal strength indication (RSSI) method to determine the estimated distances of each location of sensor node [6], then uses a least-squares trilateration (LST) algorithm [7] to calculate the concerned node location. The most attractive advantage of indoor LAS is that it can substitute the global positioning system (GPS) for indoor positioning applications. This is because the GPS can not receive / transmit any satellite signals indoor because of many obstacles of building. The LAS has been used in many university or institute for the research purpose; this LAS is named *i-Tracer* system produced by *Fontal Technology Inc. Taiwan* [2]. By using the *i-Tracer,* many indoor LAS applications can be developed. In this paper, the patient's position monitoring methodology is proposed.

ANFIS has been proposed many years ago and widely used in research works [2]. It reveals an efficient learning network and its applications can be found in many articles. ANFIS is a suitable off-line hybrid learning network to serve as a basis for constructing a set of fuzzy if-then rules with appropriate membership functions to generate the stipulated fuzzy associated memory (FAM) input-output pairs and fuzzy rules. Then a fuzzy inference system (FIS) matrix can be achieved. By using this FIS matrix, a reliable controller can be designed. The *MATLAB*™ has provided very useful, powerful and friendly tools for engineers to design this ANFIS controller [8].

The research topics about patient' applications have been grown recently due to the great interest among people. Some topics have been developed such as health insurance system [9], advisory decision monitoring patients system [10] or remote health monitoring system [11], recognizing of human activity system [12] and nursing homes for elderly patients system [13]. The above mentioned articles have demonstrated good and useful benefits for patients' applications. Based on this reason, LAS is used and is combined with ANFIS to develop the position monitoring system for some special requirements of patients.

In this paper, the LAS will be demonstrated first then it will be applied for the patient's position monitoring. In empirical test, good performance of position monitoring of LAS are possessed. This method also takes great benefits for reducing many burdens of nurses or doctors for caring the patients' positions.

2 Anfis-based Location Awareness System (LAS)

An indoor LAS considers only indoor environments such as inside a building. The location of users or their devices called 'Tag' can be determined by the server of

LAS. From this definition, LAS can work all the day and can offer useful position information of the 'Tag' of LAS. The LAS used in this paper is called *i-Tracer*, it can calculate the signal path loss due to signal propagation attenuation or RSSI. For example, if there are three location nodes, three path losses can be denoted as pl_1, pl_2 and pl_3, the location of tag 'A' can be calculated by means of Least-Squares Trilateration (LST) methodology as shown in Fig. 1. The values of path losses or RSSI will be varied according to the distances between the Tags and location nodes. Near distance will induce larger RSSI value and vice versa. All data can be stored in the database of LAS by Structured Query Language (SQL) and Java language programs. The LAS is constructed by 3 servers to manage the data processing, positioning and monitoring.

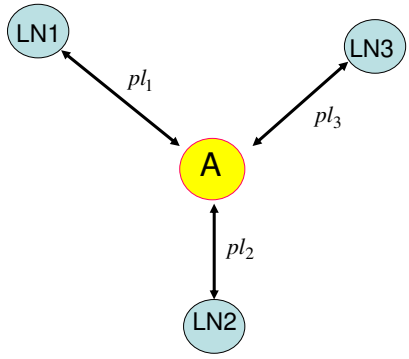


Figure 1
The concept diagram of least-squares trilateration (LST)

The ANFIS is a useful tool developed by intelligent learning algorithm to identify the membership function's parameters to generate reliable fuzzy inference systems (FIS). It uses the method of least-squares methods to train FIS membership function parameters from a given training input/output data set. The principle of ANFIS is described below [2]:

$$R_i : If\ x_1\ is\ A_{i1}\ \ ...and\ x_n\ is\ A_{in}\ then\ u_i = p_{i1}x_1 + .. + p_{in}x_n + r_i \qquad (1)$$

where R_i denotes the *i*th fuzzy rules, *i*=1, 2,..,*j*; A_{ik} is the fuzzy set in the antecedent associated with the *k*th input variable at the *i*th fuzzy rule; and $p_{i1},..., p_{in}$, r_i are the fuzzy consequent parameters.

Based on the *weighted averaged method* of defuzzification, the output u can be calculated as

$$u = \frac{w_1}{w_1 + .. + w_j} u_1 + .. + \frac{w_j}{w_1 + .. + w_j} u_j = \overline{w}_1 u_1 + .. + \overline{w}_j u_j \tag{2}$$

where w_i is the ith node output firing strength of the ith rule; and $\overline{w}_1 = \frac{w_1}{w_1 + .. + w_j}, .., \overline{w}_j = \frac{w_j}{w_1 + .. + w_j}$.

Because the fuzzy inference system is a Takagi-Sugeno (T-S) type, i.e. $u_i = p_{i1} x_1 + .. + p_{in} x_n + r_i$, equation (2) can be rewritten as

$$\begin{aligned} u &= \overline{w}_1 u_1 + .. + \overline{w}_j u_j \\ &= (\overline{w}_1 x_1) p_{11} + .. + (\overline{w}_1 x_n) p_{1n} + (\overline{w}_1) r_1 \\ &\quad + \\ &\quad \vdots \\ &+ (\overline{w}_j x_1) p_{j1} + .. + (\overline{w}_j x_n) p_{jn} + (\overline{w}_j) r_j . \end{aligned} \tag{3}$$

A neural network structure of ANFIS is shown in Fig. 2. There are five layers which includes input layer, input membership function layer, rule layer, output membership function layer and output layer. All the parameters of layers can be trained by typing command in Matlab™ window. By means of ANFIS-based control methodology, the designed methodology can be carried out. The overall design conceptual diagram is shown in Fig. 3.

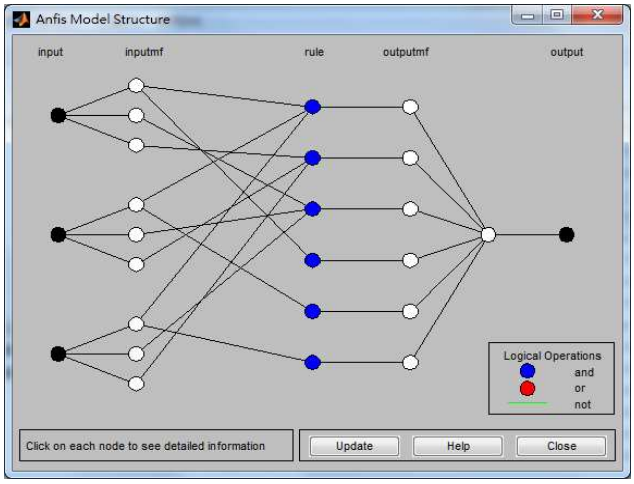


Figure 2
Diagram of ANFIS architecture

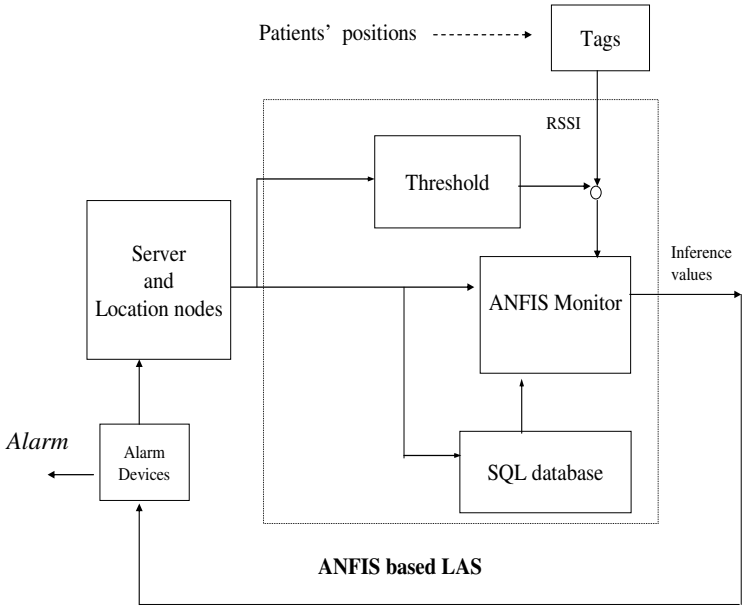


Figure 3
Conceptual diagram of ANFIS based LAS for patient's position monitoring

3 Experiment and Simulation Results

3.1 Experiments Implementation

For the experimental tests, the aforementioned values of path loss or RSSI will be varied according to the distance between the location nodes and tags. One of the LAS experimental results diagram is shown in Fig. 4, the tag is worn on the puppet neck. The LAS is constructed by 3 location nodes to manage processing data and monitoring tag's location. The experimental test is carried out by using three location nodes and one tag. At first, the gateway should be connected successfully; then the message diagram of LAS is shown in Fig. 5. It shows that one gateway and three location nodes are setup successfully. One of experimental tests of Tag of '9043' is located near location node 3 which is shown in Fig. 6. This diagram is corresponding to Fig. 4 (a). The tag number will be displayed beside the nearest location node. In this program, the actual analogue to digital (A/D) value will be normalized as hexadecimal values from 0x00 to 0xFF. Finally, data of the LAS experiment results diagram is shown in Fig. 7, the tag of number '9043' is near the

location node 3. By using *i-Tracer* graphic user interface (GUI), the data can be appeared in the screen and stored in database of SQL. From Fig. 7, for example, the first socket data begins with label of #31 is useful; the meaning of the first string received in the first line of Fig. 7 is explained by marked in bold as follows:

#31;9043;**04**;0008;2493;**E4**;0008;0000;**B4**;0008;0001;**96**;0008;124A;**90**;2D. (4)

It means that the Tag number is '9043' and it has received four hexadecimal values of path loss or RSSI, these are E4, B4, 96 and 90, respectively. All these data will be stored in database by SQL program and can be fed to ANFIS system to construct a patient's position monitoring system. The SQL database is installed successfully and can be started to work such as in Fig. 8.

3.2 Simulation Demonstration

For the simulation, the position data should be got in advance. So LAS must be set up appropriately at first. From the concept of Fig. 1 to Fig. 3, the path loss values or RSSI between tag and location nodes will be stored in LAS server. All necessary data can be achieved by using SQL language program such as in Fig. 8. In simulation, one tag is extended to three tags which are used to verify the proposed monitoring methodology. For simplicity, one case is explained as follows. One of the simulation results are shown in Fig. 9. The inference ANFIS output value is normalized from 0 to 1. If the patient's position is away from the suitable position, the ANFIS value will be reduced down; on the contrary, if the patient's position is near the indicated position, the ANFIS value will be high. If the value is lower than the threshold value, then the alarm signals such as sound, image or light will be issued. In this condition, nurses or doctors can take care or pay more attentions to that patient. For example, from Fig. 9 (a) of this simulation, patient of tag 1 has left his room; the ANFIS will alarm the signal because it has low fuzzy inference value HC= 0.0167 which is lower than the threshold value of 0.5. On the other hand, when all patients are in rooms the result is shown in Fig. 9 (b), it has high fuzzy inference value HC= 0.981; the ANFIS will not alarm any signals.

The ANFIS-based LAS for patient's position monitoring in this paper just illustrate an example of patient's position monitoring. There are many other applications can be developed by this LAS such as smart light systems, intelligent indoor navigation systems and firefighting rescue systems, etc. From this example, it demonstrates that the experimental and ANFIS-based LAS simulations have been successfully established; meanwhile, the good patient's position monitoring performance is achieved. Normally, three location nodes are necessary for the LAS, but if the accuracy of system is an important issue, more location nodes can be used to increase the accuracy of this system.

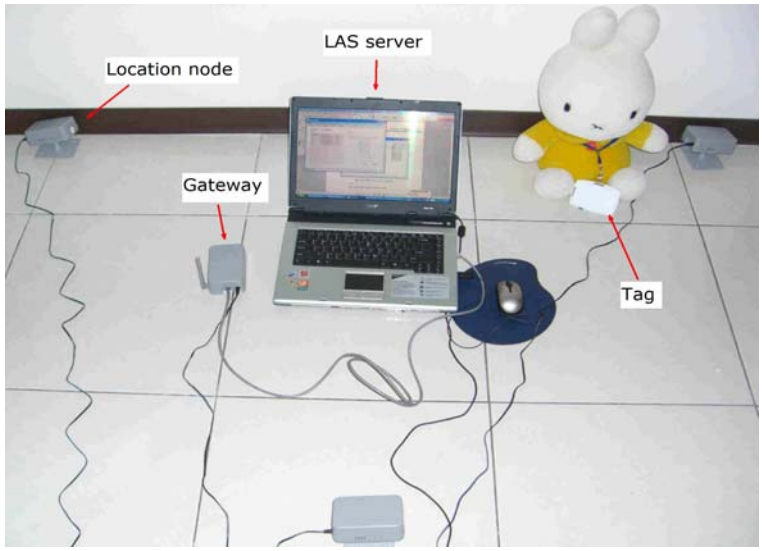


Figure 4(a)
The architecture photograph of LAS

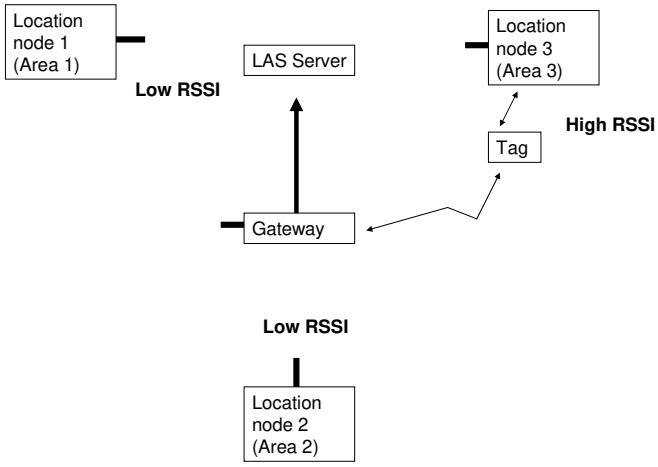


Figure 4(b)
The concept diagram of LAS

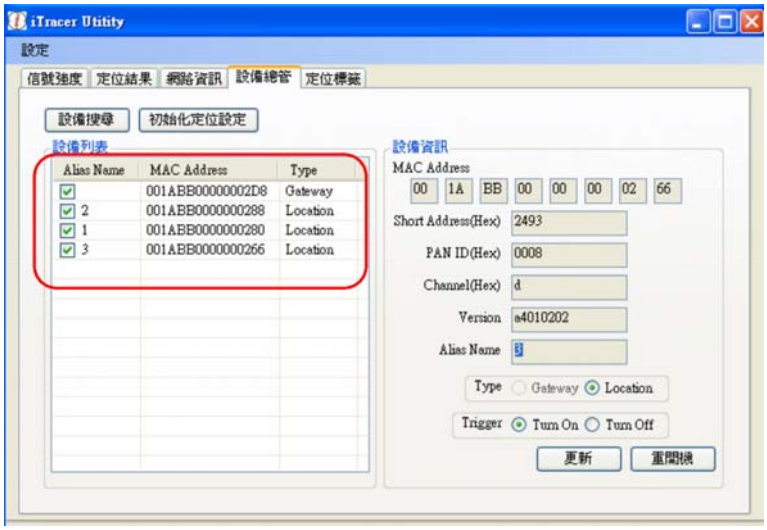


Figure 5
The successful connection diagram of LAS

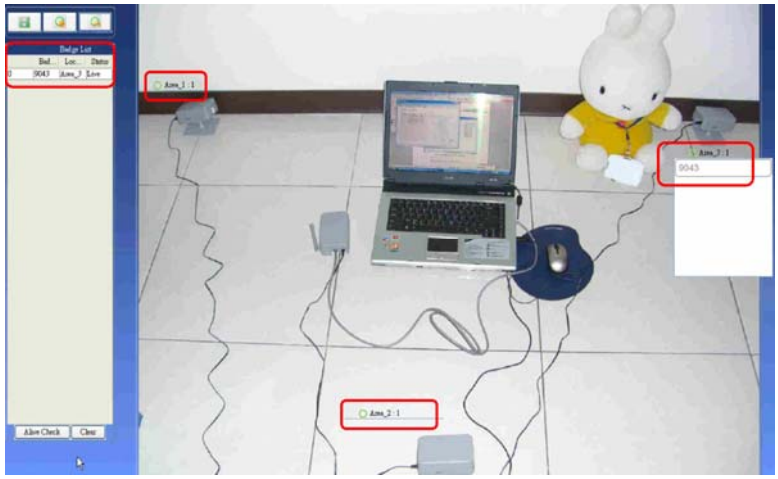


Figure 6
The diagram of LAS for Tag of '9043' is located in location node 3

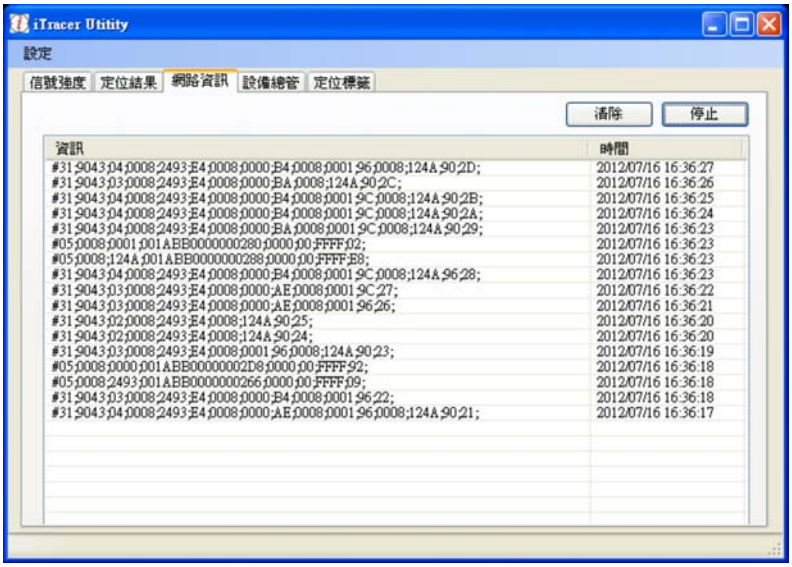


Figure 7
Diagram of implementation of *i-Tracer* graphic user interface (GUI)

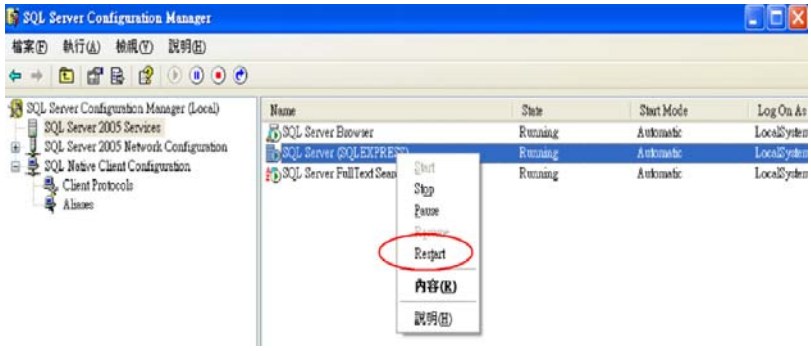


Figure 8
Diagram of SQL database working for LAS

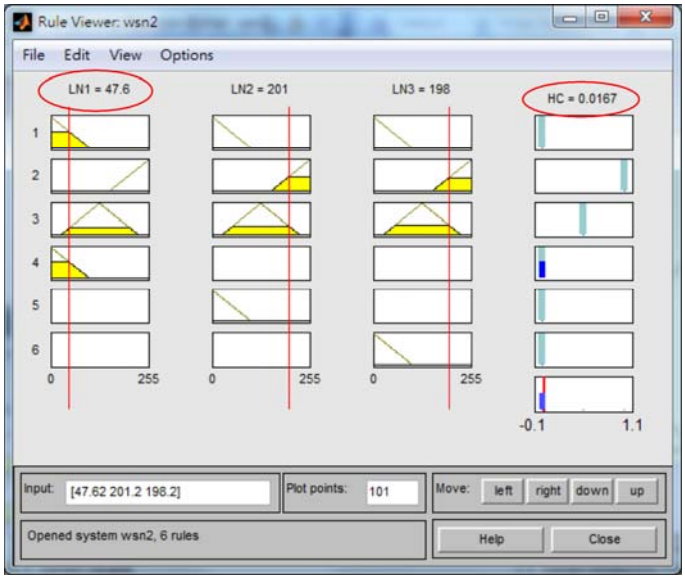


Figure 9 (a)
The ANFIS inference result diagram (patient 1 has left the room)

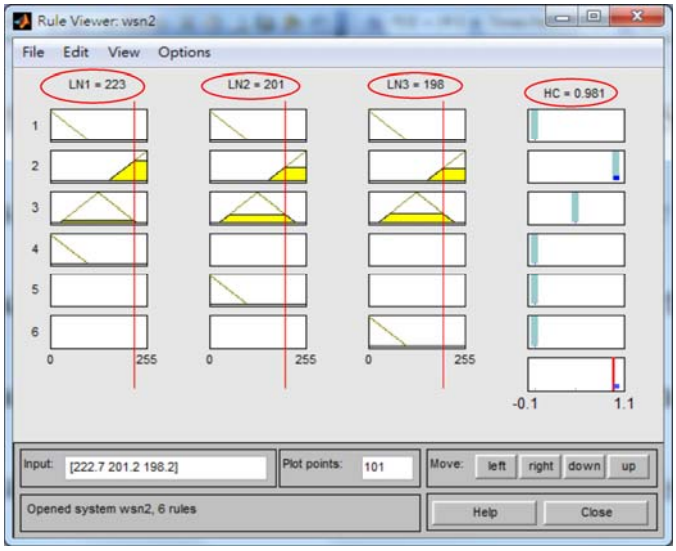


Figure 9 (b)
The ANFIS inference result diagram (all patients are in rooms)

Conclusion

In this paper, the design method for the application of patient's position monitoring by using the adaptive network based fuzzy inference system (ANFIS) based location awareness system (LAS) is proposed. This study has successfully demonstrated the application of the LAS to monitor the location of patients. The physical implementations and simulations are also successfully demonstrated, which show this system has possessed the good performance of ANFIS-based LAS for the position monitoring of patients.

References

[1] i-Tracer User Guide, *Fontal Technology Inc., Taiwan*, 2013 (http://www.fontaltech.com.tw)

[2] J. S. R. Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference System," *IEEE Transactions on Systems, Man and Cybernetics,* Vol. 23, pp. 665-685, 1993

[3] P. Baronti, P. Pillai, V. W. C. Chook, S. Chessa, A. Gotta, Y. F. Hu, "Wireless Sensor Networks: A Survey on the State of the Art and the 802.15.4 and ZigBee Standards," *Computer Communications,* Vol. 30, pp. 1655-1695, 2007

[4] Y. J. Mon, C. M. Lin, I. J. Rudas, "Wireless Sensor Network (WSN) Control for Indoor Temperature Monitoring," *Acta Polytechnica Hungarica,* Vol. 9, No. 6, pp. 17-28, 2012

[5] Y. J. Mon, C. M. Lin, I. J. Rudas, "ANFIS-based Wireless Sensor Network (WSN) Applications for Air Conditioner Control," *Acta Polytechnica Hungarica,* Vol. 10, No. 3, pp. 5-16, 2013

[6] Y. J. Mon, "The Distance Measurement by Using RSSI of Wireless Sensor Network," *International Journal of Computational Engineering Research,* Vol. 3, No. 4, pp. 110-112, 2013

[7] K. Lu, X. Xiang, D. Zhang, R. Mao and Y. Feng, "Localization Algorithm Based on Maximum a Posteriori in Wireless Sensor Networks," *International Journal of Distributed Sensor Networks*, Article ID 260302, 2012 (DOI:10.1155/2012/260302)

[8] MATLAB User Guide, *Mathworks Inc.*, 2012 (http://www.mathworks.com)

[9] M. F. Abbod, D. G. Keyserlingk, D. A. Linkens, M, Mahfouf, "Survey of Utilisation of Fuzzy Technology in Medicine and Healthcare," *Fuzzy Sets and Systems,* Vol. 120, pp. 331-349, 2001

[10] J. Bajo, J. F. Paz, Y. Paz, J. M. Corchado, "Integrating Case-based Planning and RPTW Neural Networks to Construct an Intelligent Environment for Health Care," *Expert Systems with Applications,* Vol. 36, pp. 5844-5858, 2009

[11] J. C. Huang, "Remote Health Monitoring Adoption Model Based on Artificial Neural Networks," *Expert Systems with Applications,* Vol. 37, pp. 307-314, 2010

[12] Y. J. Mon, "Health Care Activity Map Designed by Fuzzy Neural Network," *Advanced Science, Engineering and Medicine*, Vol. 4, pp. 355-359, 2012

[13] M. P. Rajasekaran, S. Radhakrishnan, P. Subbaraj, "Sensor Grid Applications in Patient Monitoring," *Future Generation Computer Systems,* Vol. 26, pp. 569-575, 2010

Gain-scheduled Adaptive Observer for Induction Motors: An LMI Approach

Fethi Farhani[1], Abderrahmen Zaafouri[2], Abdelkader Chaari[3]

University of Tunis, Unit C3S, Higher School of Sciences and Techniques of Tunis (ESSTT), 5 Av. Taha Hussein, BP 56, 1008 Tunis, Tunisia;
[1]E-mail: fethi.farhani@issatkr.rnu.tn
[2]E-mail: abderrahmen.zaafouri@isetr.rnu.tn
[3]E-mail: assil.chaari@esstt.rnu.tn

Abstract: This paper presents a new adaptive rotor flux observer with both speed and rotor and stator resistance identification. The proposed solution , which is based on a full-order Luenberger observer, guarantees a perfect reconstruction of state variables while reducing the drive system complexity. By using the Lyapunov's direct method, the gain-scheduled of the suggested observer can be achieved by solving four bilinear matrix inequalities using the LMI Toolbox of MATLAB. To overcome the problem of parametric variation of induction machine, a tracking-parameter mechanism is used. To validate the effectiveness and robustness of the proposed solution, some simulation results are provided.

Keywords: Asymptotic stability; Gain-scheduled; linear matrix inequalities; Pole Placement; State estimation

1 Introduction

The three-phase squirrel-cage induction machine driven by static converters is the most used in high-performance industrial applications. It currently occupies an important place in rotating electrical machines market thanks to its robustness, simplicity, low-cost manufacturing and high specific power.

Nowadays, as a consequence of the important progress realized in the technology of microcontrollers and electronic power switches, it is possible to realize variable speed drives using vector control based on complex algorithms while taking into account the difficulties related to the nonlinearities of the induction machine. The vector control introduced by Blaschke [1] has improved the performances of the induction machine so that it provided performances similar to the DC machine. In the literature, there are two main techniques of field-oriented control: direct and indirect orientation [2]-[5]. Both techniques require an online recognition of the state variables. This involves finding software solutions to reconstruct the

immeasurable variables such as rotor or stator flux and even measurable variables such as rotation speed to improve the reliability and robustness of the control schemas.

In the literature, several solutions based on the model of induction machines provide an estimate of the flux vector [6]-[8].These techniques are highly sensitive to parametric variations [9]. This imposes the search for methods to reduce the influence of these disturbances on the stability and efficiency of both state estimation and control algorithm of induction machines. In this context we propose a solution derived from the theory of Lyapunov which ensures the online adaptation of the rotor and stator resistances of the induction machine.

Direct measurement of mechanical speed by using an optical tachometer or a shaft-mounted encoder greatly increases the cost and reduces the reliability. However, even if a speed sensor is used, we can envisage a control strategy capable of tolerating sensor failure. In the last two decades, several techniques have been employed for speed sensor emulation. Rotor speed estimation techniques can be mainly classified into three methods; the first technique is based on the injection of high frequency signals to track the mechanical speed [10], [11]. The second technique is based on Kalman filter for rotor speed tracking [12]. The third technique is based on adaptive mechanisms hence its name [13].

The present article is organized as follows: In section 2, the model of induction motor will be defined. Section 3 is devoted to developing the adaptive full-order observer on the one hand and to present online parameters and speed tracking mechanisms on the other. Improving performance and robustness of the proposed observer is established in section 4. In section 5 we present the online procedure for efficiency optimization of induction motor. The validation of the proposed solution as well as the discussion of the results are carried out in section 6. Finally, a conclusion will be provided.

2 Induction Machine Modeling

The dynamics of the induction machine is defined by the following equation

$$\begin{cases} \dot{x} = \left(A + \omega_r A_\omega \right) x + B v_s \\ i_s = Cx \end{cases} \tag{1}$$

The time derivative of stator flux in a stationary reference frame is given as follows

$$\dot{\lambda}_s = v_s - R_s i_s \tag{2}$$

We deduce the rotor flux expressions (3) as a function of the stator flux and current

$$\lambda_r = D\left[\lambda_s \quad i_s\right]^T = C_1 x \tag{3}$$

In the equations (1), (2) and (3), the following symbols are used:

$$x = \left[i_s \quad \lambda_r\right]^T ; A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}; A_\omega = \begin{bmatrix} 0_{2x2} & A_{\omega 1} \\ 0_{2x2} & A_{\omega 2} \end{bmatrix}; B = \left[B_1 \quad 0_{2x2}\right]^T ;$$

$$C = \left[I \quad 0_{2x2}\right]; D = \left[M / L_s\left(1-\sigma\right) \quad \sigma M / L_s\left(\sigma - 1\right)\right]; C_1 = \left[0_{2x2} \quad I\right];$$

$$A_{11} = -\left(R_r\left(1-\sigma\right)/\sigma L_r + R_s / \sigma L_s\right) I ; A_{12} = \left(R_r / \varepsilon L_r\right) I ; \varepsilon = \sigma L_r L_s / M$$

$$A_{21} = \left(M R_r / L_r\right) I ; A_{22} = \left(-R_r / L_r\right) I ; A_{\omega 1} = -\left(1/\varepsilon\right) A_{\omega 2} = -\left(1/\varepsilon\right) J ;$$

$$B_1 = \left(1/\sigma L_s\right) I ; I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

R_s, R_r : Stator and Rotor resistance

L_s, L_r : Stator and Rotor inductance

M : Mutual inductance

λ_s, λ_r : Space vector of stator and rotor flux

i_s : Space vector of the stator currents

$\sigma = 1 - \left(M^2 / L_r L_s\right)$: Total leakage factor

ω_r : Angular rotor speed

p : Number of pole pairs

Equation (4) defines the electromagnetic torque in the stator reference frame.

$$Ce = p\left(3M / 2L_r\right)\left(\lambda_{r\alpha} i_{s\beta} - \lambda_{r\beta} i_{s\alpha}\right) \tag{4}$$

3 The adaptive Flux Observer

To reconstruct the state variables and to estimate the mechanical speed of the induction machine, a full-order observer is proposed and expressed in its simplest form as shown in the differential equation (5). The observer gain ***H*** appears to be obvious. The symbol ^ above a symbol denotes an estimate value of a parameter or a state variable.

$$\dot{\hat{x}} = \left(\hat{A} + \hat{\omega}_r A_\omega\right)\hat{x} + \hat{B}v_s + H\left(\hat{i}_s - i_s\right) \tag{5}$$

Where $\hat{A} = A + \Delta A$, $\hat{B} = B + \Delta B$ and $\hat{C} = C + \Delta C$

In this article, we focus only on the resistive uncertainty. The estimation error $e = x - \hat{x}$ is defined as the difference between real and estimated vectors. The dynamic of estimation error of the rotor flux is described by the following equation (6)

$$\dot{e} = \left(A + HC + \omega_r A_\omega\right)e + \Delta\omega_r A_\omega \hat{x} + \Delta R_s A_{Rs}\hat{x} \tag{6}$$

Where

$$\Delta\omega_r = \hat{\omega}_r - \omega_r \text{ , } \Delta A = \left(\hat{R}_s - R_s\right)A_{Rs} = \Delta R_s A_{Rs} \text{ and } A_{Rs} = \begin{bmatrix} -\left(1/\sigma L_s\right)I & 0_{2x2} \\ 0_{2x2} & 0_{2x2} \end{bmatrix}$$

To ensure the asymptotic stability of the state estimation error, we consider the following quadratic Lyapunov function

$$V\left(e, \Delta\omega_r, \Delta R_s\right) = e^T Pe + \left(2\Delta\omega_r{}^2 / \mu_1\right) + \left(2\Delta R_s{}^2 / \mu_2\right) \tag{7}$$

Where P is a positive-definite symmetric matrix, μ_1 and μ_2 two positive design constants.

The time derivative of V is:

$$\dot{V}\left(e, \Delta\omega_r, \Delta R_s\right) = \dot{e}^T Pe + e^T P\dot{e} + \frac{2}{\mu_1}\left(\Delta\omega_r\right)\left(\frac{d\Delta\omega_r}{dt}\right) + \frac{2}{\mu_2}\left(\Delta R_s\right)\frac{d\Delta R_s}{dt} \tag{8}$$

After some calculation, equation (8) becomes

$$\dot{V}\left(e, \Delta\omega_r, \Delta R_s\right) = e^T\left(\left(A + HC + \omega_r A_\omega\right)^T P + P\left(A + HC + \omega_r A_\omega\right)\right)e + V_1 + V_2 \tag{9}$$

Where V_1 and V_2 are given by the following equations

$$V_1 = \Delta\omega_r\left(\left(\hat{x}^T A_\omega^T Pe + e^T PA_\omega\hat{x}\right) + \frac{2}{\mu_1}\left(\frac{d\left(\hat{\omega}_r - \omega_r\right)}{dt}\right)\right) \tag{10}$$

$$V_1 = \Delta R_s\left(\left(\hat{x}^T A_{Rs}^T Pe + e^T PA_{Rs}\hat{x}\right) + \frac{2}{\mu_2}\frac{d\Delta\left(\hat{R}_s - R_s\right)}{dt}\right) \tag{11}$$

We consider the variation dynamic of the speed as slower than the error estimation dynamic on the mechanical speed ($\dot{\omega}_r \approx 0$). The same conditions apply to the dynamic of stator resistance variation and its error estimation dynamic ($\dot{R}_s \approx 0$).

By nullifying the sum of V_1 and V_2, the relationship between the state error estimation and the dynamics of both speed and stator resistance estimation can be expressed as follows

$$\frac{2}{\mu_1}\frac{d\hat{\omega}_r}{dt} = -\hat{x}^T A_\omega^T Pe - e^T PA_\omega \hat{x} \tag{12}$$

$$\frac{2}{\mu_2}\frac{d\hat{R}_s}{dt} = -\hat{x}^T A_{Rs}^T Pe - e^T PA_{Rs} \hat{x} \tag{13}$$

The first and second terms on the right-hand side of both equations (12) and (13) are scalars; the adaptive law of estimation of the rotor speed and stator resistance can be set in the following equation

$$\begin{cases} \dfrac{d\hat{\omega}_r}{dt} = -\mu_1 e^T PA_\omega \hat{x} \\ \dfrac{d\hat{R}_s}{dt} = -\mu_2 e^T PA_{Rs} \hat{x} \end{cases} \tag{14}$$

If (14) holds and the gain matrix, H, is set in order that the time derivative of Lyapunov equation (9) is nonpositive, this means that the estimation error converges to zero and the observer tracks the real system. In the proposed solution, a PI regulator is used instead of the integral controller (14) to improve the performances of the speed and stator resistance estimators. The adaptation lows of the rotor speed and stator resistance are defined respectively by following equations (15) and (16)

$$\hat{\omega}_r = K_{pv}\left(e^T PA_\omega \hat{x}\right) + K_{iv}\int\left(e^T PA_\omega \hat{x}\right)dt \tag{15}$$

$$\hat{R}_s = K_{pRs}\left(e^T PA_{Rs} \hat{x}\right) + K_{iRs}\int\left(e^T PA_{Rs} \hat{x}\right)dt \tag{16}$$

Where K_{pv}, K_{iv}, K_{pRs} and K_{iRs} are positive design constants

Based on the simplified thermic model of induction machine, and supposing that rotor and stator coils have almost the same temperature, an adaptive mechanism of rotor resistance will be described as follows

$$\hat{R}_r = R_r\left(1 + \left(\alpha_{Rr} / \alpha_{Rs}\right)\left(\left(\hat{R}_s / R_s\right) - 1\right)\right) \tag{17}$$

α_{Rs}: Temperature coefficient of stator resistance.

α_{Rr}: Temperature coefficient of rotor resistance.

4 Observer Gain Calculation: Pole Placement in LMI Region

To enhance the dynamic behavior of the proposed flux observer, a robust pole placement algorithm is proposed. The proposed solution uses LMI (Linear Matrix Inequalities) technique. It is suggested to ensure the matrix ($A+\omega_r A_\omega+HC$) pole placement in specific LMI-regions. The latter is defined by the intersection between a circle centered at center (0,0) with radius (r) and the left half plan limited by a vertical line with the abscissa ($-h$) where h is positive constant (Figure 1). According to [14] and [15], the LMI formulation of the proposed set up is given by the following equation

$$\begin{cases} \begin{bmatrix} -rP & A_{e\omega}^T P + C^T H^T P \\ PA_{e\omega} + PHC & -rP \end{bmatrix} < 0 \\ PA_{e\omega} + A_{e\omega}^T P + PHC + C^T H^T P + 2hP < 0 \end{cases} \tag{18}$$

With

$P = P^T > 0$ And $A_{e\omega} = A + \omega_r A_\omega$

The bilinear matrix inequality (BMI) (18) can be transformed into LMIs with a change of variables $R = PH$ as follows:

$$\begin{cases} \begin{bmatrix} -rP & A_{e\omega}^T P + C^T R^T \\ PA_{e\omega} + RC & -rP \end{bmatrix} < 0 \\ PA_{e\omega} + A_{e\omega}^T P + RC + C^T R^T + 2hP < 0 \end{cases} \tag{19}$$

The matrix $A_{e\omega}$ depends affinely on the rotation speed $\omega_r \in \begin{bmatrix} \omega_{r1} & \omega_{r2} \end{bmatrix}$; where $A_{e\omega i}$ denotes the parameter values of $A_{e\omega}$ at the vertices ω_{ri} of the parameter polytope. According [9] it is then possible to obtain the observer gain $\boldsymbol{H}$ if and only if there exist real positive matrices $P = P^T$, R_1 and R_2 simultaneously satisfy the four following matrices inequalities:

$$\begin{cases} \begin{bmatrix} -rP & A_{e\omega i}^T P + C^T R_i^T \\ PA_{e\omega i} + R_i C & -rP \end{bmatrix} < 0 \\ PA_{e\omega i} + A_{e\omega i}^T P + R_i C + C^T R_i^T + 2hP < 0 \end{cases} \qquad i \in \{1,2\} \tag{20}$$

The decision variables P, R_1 and R_2 can be synthetized numerically under LMI constraints of (20) by using the LMI Control Toolbox of Matlab. As a result, the observer sub-gain H_i can be calculated by the following equation

$$H_i = P^{-1} R_i \qquad i \in \{1,2\} \tag{21}$$

The observer gain H for a given $\omega_r \in [\omega_{r1} \quad \omega_{r2}]$ can be obtained by the interpolation between the two sub-gains H_1 and H_2, as given in the following equation

$$H(\omega_r) = \frac{H_1(\omega_{r2} - \omega_r) + H_2(\omega_r - \omega_{r1})}{\omega_{r2} - \omega_{r1}} \tag{22}$$

Finally, the block diagram of the proposed Adaptive state observer of induction machine is shown in Figure 2.

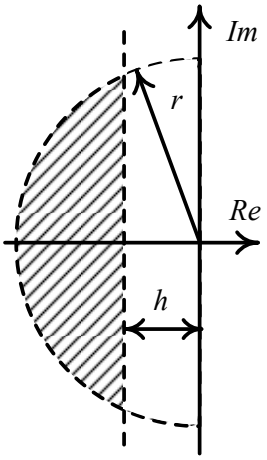


Figure 1
Proposed region

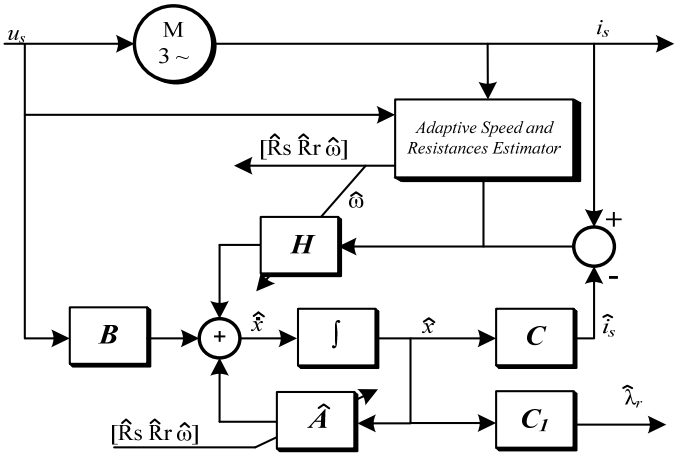


Figure 2
Adaptive state observer of induction machine

5 Model-based Efficiency Optimization

For maintenance reasons, some users prefer the use of electric motors of the same size. Almost half of the electric under-loaded engines operate at less than 40% of their rated loads [16]. This requires a revision of the strategy control in order to optimize the electrical energy losses. An energy analysis shows that the iron and copper losses represent over 80% of all losses in an induction motor [17]. To ensure the minimization of energy losses, in [18] we have proposed an adaptive flux control based on the research of optimal flux which corresponds to an operating speed and a load torque (23).

$$\lambda_{rd} = M \sqrt[4]{\frac{R_q}{R_d}} \sqrt{kCe} \tag{23}$$

Where

$$R_d = R_s + \left(R_s + 1\right)\left(\omega_s L_s\right)^2 / R_c$$

$$R_q = R_s + \left(R_s R_c + 1\right) + \left(\left(\omega_s L_s \sigma\right)^2 / R_c\right) + R_r \left(M / L_r\right)^2$$

$$k = \left(2L_r\right) / \left(3pM^2\right)$$

6 Results

To verify the proposed solution, a discrete Simulink model in Matlab® is built based on the control block diagram shown in Figure 3. The parameters of induction motor are shown in Table 1. The electric power devices including the voltage source inverter, the three-phase induction motor are emulated by Powergui. The adaptive observer and the vector control are programmed with a 10 µs sample time. Figure 4 shows the system response to a step speed (500 rpm) under a load torque of 20 Nm, the motor takes 50 ms to reach its steady state. Figure 5 shows the developed electromagnetic torque during the starting phase as well as the stationary phase. The output signal of speed estimator block is used as a feedback for the speed closed loop regulation based on an Anti-Windup PI controller. To verify disturbance rejection on speed control, the load torque is increased by 50% at t=1.5 s (Figure 5). The proposed solution shows a good performance on the speed track (Figure 4), and disturbance rejection in the steady-state. Indeed, the estimated speed converges correctly towards the actual speed.

The three-phase inverter feeds the induction machine based on the control algorithm. In the startup phase, the control system ensures the magnetization of induction motor based on a specific magnetization algorithm. When the rotor flux

reaches the minimum level, the control system switches to the DFOC algorithm. The first transition lasts 10 ms; the second 240 ms, as illustrated in Figure 6.

The output signal of the proposed rotor flux observer is re-injected into the input of rotor flux regulation block based on PI controller. The optimal rotor flux magnitude is computed by the Energy Optimization Algorithm block (EOA). At t=1 s, the drive system switches from a classical DFOC to a DFOC under optimal rotor flux; Figure 6 shows that the magnitude of reference rotor flux has decreased to 84% of its rated value. Figure 6 also shows the convergence of the estimated rotor flux to the actual value; a good tracking of the reference is successfully guaranteed by the controller. Due to the sensitivity of the EOA to the stator and rotor resistance variation, the estimator of stator and rotor resistances provides the actual value to the optimization algorithm. To verify the performance of the parameters' tracking system, the actual values of stator and rotor resistances are increased by 20%. Figure 7 and Figure 8 show that the estimated values of rotor and stator resistance converge respectively to their real values.

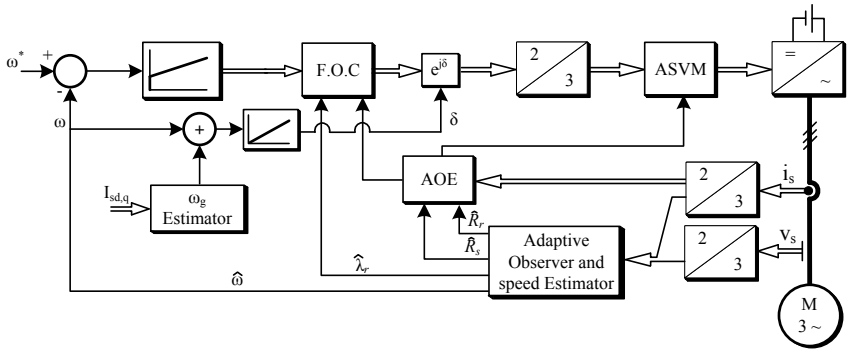


Figure 3

Block diagram of a speed Sensorless control system based on adaptive observer (**AOE**: **E**nergy **O**ptimization **A**lgorithm)

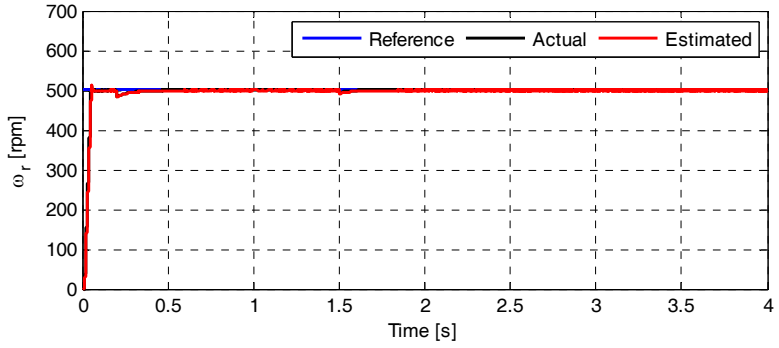


Figure 4

Simulated speed using the proposed speed adaptive scheme

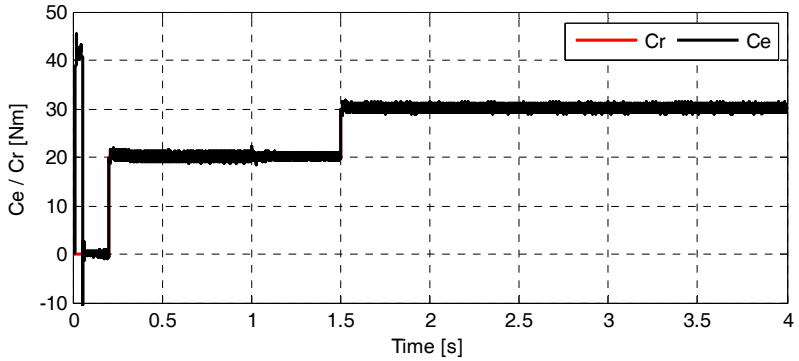


Figure 5
Torque response for step varying of load torque
(Cr: load torque, Ce: electromagnetic torque)

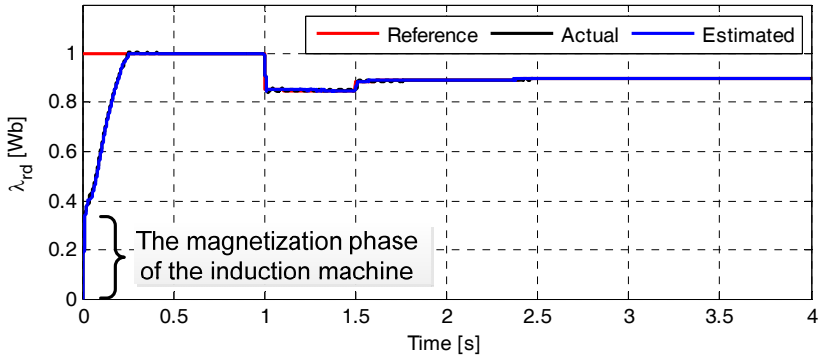


Figure 6
The curve of real and estimated rotor flux

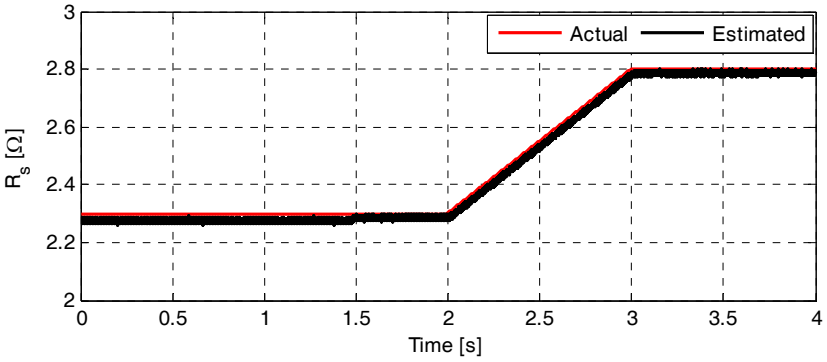


Figure 7
The curve of the stator resistance variation

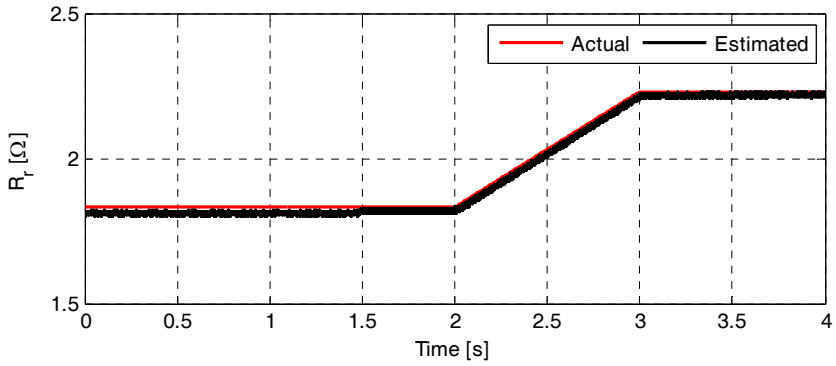


Figure 8
The curve of the rotor resistance variation

Table 1
Induction machine parameters

Symbol	Quantity	Rating values
R_s	Stator resistance	2.3 Ω
R_r	Rotor resistance	1.83 Ω
L_s	Stator inductance	0.261 H
L_r	Rotor inductance	0.261 H
M	Mutual inductance	0.245 H
σ	Leakage factor	0.134
jm	Moment of inertia	0.03 Kgm²
f	Friction coefficient	0.001
U_n	Rated voltage	380 V
I_n	Rated current	15.1 A
P_n	Rated power	7 KW
p	Number of pole pair	-

Conclusion

In this paper, a new Sensorless induction machine drive has been proposed and tested with simulation in the Matlab/Simulink environment. The proposed concept is based on DFOC and a robust flux observer. The observer gain is deduced from Lyapunov theory. To enhance the dynamic and static performances of the proposed observer, the pole placement technique has been used. The observer gain is obtained from solving linear matrix inequalities by using the LMI Toolbox of MATLAB®. Furthermore, the tracking of the actual value of rotor speed as well as stator and rotor resistances is ensured by a PI adaptive law deduced from Lyapunov theory. Along with the DFOC algorithm, an optimization strategy of the energy losses of the induction machine has been introduced.The simulation results show the high static and dynamic perfermances.

Acknowledgement

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions to improve the original manuscript.

References

[1] F. Blaschke, "The Principle of Field-Orientation as Applied to the New 'Transvector' Closed Loop Control System for Rotating-Field Machines," Siemens-Review, Vol. 34, No. 5, pp. 217-220, 1972

[2] T. Jebali, M. Jemli, M. Boussak, M. Gossa, M. B. Kamoun, "Dspace-based Experimental Results of Indirect Field oriented Control (IFOC) PWM VSI Fed Induction Motor," in IEEE International Conference on Industrial Technology, Hammamet, Tunisia, 2004, Vol. 2, pp. 569-573

[3] A. Laoufi, A. Hazzab, I. K. Bousserhane, M. Rahli, "Direct Field-oriented Control using Backstepping Technique for Induction Motor Speed Control," in Information and Communication Technologies. 2nd, Damascus, Syria, 2006, Vol. 1, pp. 1422-1427

[4] B. B. S, V. A. M, "Speed-Sensorless, Adjustable-Speed Induction Motor Drive Based on dc Link Measurement," International Journal of Physical Sciences, Vol. 4, No. 4, pp. 221-232, Apr. 2009

[5] R. Sadouni, A. Meroufel, "Indirect Rotor Field-oriented Control (IRFOC) of a Dual Star Induction Machine (DSIM) Using a Fuzzy Controller," Acta Polytechnica Hungarica, Vol. 9, No. 4, pp. 117-192, 2012

[6] H. Madadi Kojabadi, L. Chang, "Model Reference Adaptive System Pseudoreduced-Order Flux Observer for Very Low Speed and Zero Speed Estimation in Sensorless Induction Motor Drives," in IEEE 33rd Annual Power Electronics Specialists Conference : conference proceedings, Queensland, Australia, 2002, Vol. 1, pp. 301-305

[7] S. Iosif, P. Octavian, F. Ioan, C. Vasar, "Above Flux Estimation Issues in Induction Generators with Application at Energy Conversion Systems," Acta Polytechnica Hungarica, Vol. 3, No. 3, pp. 137-148, 2006

[8] F. R. Salmasi, T. A. Najafabadi, P. Jabehdar-Maralani, "An Adaptive Flux Observer with Online Estimation of DC-Link Voltage and Rotor Resistance for VSI-based Induction Motors," IEEE Transactions on Power Electronics, Vol. 25, No. 5, pp. 1310-1319, 2010

[9] M. B. B. Sharifian, N. Rostami, H. Hatami, "Sensorless Control of IM-based on Full-Order Luenberger Observer," in The 9th International Power and Energy Conference, Singapore, 2010, pp. 567-571

[10] C. Caruana, G. M. Asher, M. Sumner, "Performance of HF Signal Injection Techniques for Zero-Low-Frequency Vector Control of Induction Machines under Sensorless Conditions," IEEE Transactions on Industrial Electronics,, Vol. 53, No. 1, pp. 225-238, Feb. 2005

[11] R. Raute, C. Caruana, C. S. Staines, J. Cilia, N. Teske, M. Sumner, G. M. Asher, "A Review of Sensorless Control in Induction Machines Using hf Injection, Test Vectors and PWM Harmonics," in 2nd IEEE International Symposium on Sensorless Control for Electrical Drives, Birmingham, UK, 2011, pp. 47-55

[12] M. Barut, R. Demir, E. Zerdali, R. Inan, "Real-Time Implementation of Bi Input-Extended Kalman Filter-based Estimator for Speed-Sensorless Control of Induction Motors," IEEE Transactions on Industrial Electronics, Vol. 59, No. 11, pp. 4197-4206, Nov. 2012

[13] S. Wu, Y. Li, Z. Zheng, "Speed Sensorless Vector Control of Induction Motor Based on Full-Order Flux Observer," in CES/IEEE 5th International Power Electronics and Motion Control Conference : Conference Proceedings, Shanghai, China, 2006, Vol. 3, pp. 1-4

[14] M. Chilali, P. Gahinet, P. Apkarian, "Robust Pole Placement in LMI Regions," IEEE Transactions on Automatic Control, Vol. 44, No. 12, pp. 2257-2270, Dec. 1999

[15] M. Chilali, P. Gahinet, "H∞ Design with Pole Placement Constraints: an LMI Approach," IEEE Transactions on Automatic Control, Vol. 41, No. 3, pp. 358-367, Mar. 1996

[16] R. N. Hasanah, "A Contribution to Energy Saving in Induction Motors," Ph.D, EPFL, LAI, Faculté des sciences et techniques de l'ingénieur, Lausanne, 2005

[17] R. Razali, A. N Abdalla, G. Ruzlaini, C. Venkataseshaiah, "Improving Squirrel Cage Induction Motor Efficiency: Technical Review," International Journal of the Physical Sciences, Vol. 7, No. 8, pp. 1129-1140, Feb. 2012

[18] F. Farhani, A. Zaafouri, "On-Line Tuning of Efficiency Optimization of Induction Machine Drive," in 16th IEEE Mediterranean Electrotechnical Conference (MELECON), Hammamet, Tunisia, 2012, pp. 1137-1140

Reactor Failure due to Resonance in Zahedan-Iranshahr Parallel EHV Lines, Analysis and Practical Solutions

Mohammad Hamed Samimi[1], Moien Abedini[1], Amir Hossein Mostajabi[1], Davood Farokhzad[2], Hossein Ayoubzadeh[2], Amir Abbas Shayegani Akmal[1], Hossein Mohseni[1]

[1]High Voltage Research Center, School of Electrical and Computer Engineering, University of Tehran, North Kargar Avenue, IR-14395, Tehran, Iran
[2]High Voltage Transmission System and Power System Protection Office, Iran Grid Management Company, Yasemi str., Vali-asr str., 1996836111, Tehran, Iran
m.h.samimi@ut.ac.ir, m.abedini@ut.ac.ir, ah.mostajabi@ut.ac.ir, farokhzad@igmc.ir, ayoubzadeh@igmc.ir, shayegani@ut.ac.ir, mohseni@ut.ac.ir

Abstract: A reactor winding has been damaged due to induction of resonance voltages on the de-energized circuit of two parallel 230 kV shunt compensated lines between Zahedan and Iranshahr. The phenomenon is modeled in PSCAD and the safe range of shunt compensation is derived. In this study the effect of various faults on the lines, the reactor saturation and corona dissipation are considered and the results are compared to the electrostatic no-loss method results. Various ways are examined for damping the resonance condition and the best solution is chosen.

Keywords: resonance; PSCAD; shunt reactors; transmission line modeling

1 Introduction

Shunt reactors have many applications in extra high voltage transmission lines and using them has some challenges [1-7]. For example, the application of shunt reactor compensation to one or more circuits of mutually coupled multi-circuit overhead transmission lines requires special considerations beyond those ordinarily required for single-circuit lines as a result of voltages which may be coupled from one circuit to an adjacent circuit. This is particularly true when two lines are completely or partially untransposed [8, 9]. Previous measurements revealed that unusually high voltages and currents were experienced by the reactors when their associated circuit was disconnected from the system [10] which could damage them. Some papers previously developed a matrix analysis

for deriving the general curves which show the resonance voltages versus the shunt compensations [8], [11-15]. Some others have used software like EMTP for studying the resonance conditions [16, 17]. This paper presents the analysis and solutions for a similar situation using PSCAD.

A reactor winding is damaged in Zahedan-Iranshahr double-circuit 230 kV untransposed transmission line which is compensated with two 25 MVAr reactors at each end of the circuit. Figure 1 shows the circuit and the transmission line configuration. First, the Zahedan-Iranshahr line was tripped out by the distance relay because of a fault on that circuit. A voltage of about 15 percent higher than the nominal voltage was observed by the operator in this situation while the parallel line, Zahedan-Khash-Iranshahr line, was still live. By the idea that the voltage is real and because of failure of circuit breakers, the bus bar at Iranshahr substation was de-energized and the Iranshahr-Khash line was disconnected as well. Unfortunately the bottom line had remained in a resonance condition even when one part of the top line was de-energized and only one part of it was remained live, so the voltage was still observed on the first line. Therefore, the bus bar of Iranshahr substation had to be de-energized.

Since only 25 MVAr reactors are available in the short-term in Zahedan power section, damping seems to be a good solution. Using the PSCAD, various conditions are examined for damping the mentioned circuit and the best one is proposed. As well, the safe range of shunt compensation is derived for various lines statuses.

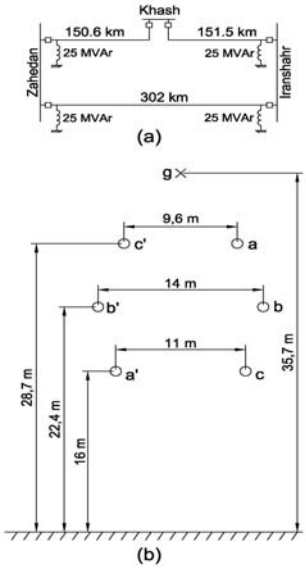


Figure 1

(a) Double-circuit 230 kV system configuration; (b) Double-circuit transmission line configuration

2 The Range of Shunt Compensation for Safe Operation

2.1 System Modeling in PSCAD

For modeling the system configuration, first we used three PSCAD simple reactors in each terminal. In this way we are able to measure the neutral currents and voltages before and after adding elements in the neutral of reactors. For modeling the transmission line configuration we used a PSCAD double-circuit transmission line tower without transposition. The dimension between conductors is the same as the one showed in Figure 1. The ACSR Canary was chosen as a phase conductor type. The sags of phase and ground wires were set to 7.5 m and 5.7 m respectively.

The soil resistivity and shunt conductance of the line have an enormous effect on the level of resulted resonance voltages. Lower shunt conductance leads to higher resonant voltages [16, 18]. A range of shunt conductance from 0.76 pS/km (non-ceramic insulator) to 6.5 nS/km (polluted glass-type insulator) in 230 kV class is reported in previous works [16]. As this transmission line is located in the desert and has ceramic type insulators, the shunt conductance of the line is set to 5 nS/km. In the case of lower values of soil resistivity, higher values of resonant voltages were computed [16]. Because of a sandy soil in the transmission line's right of the way, a value of 500 Ω.m was selected as the soil resistivity.

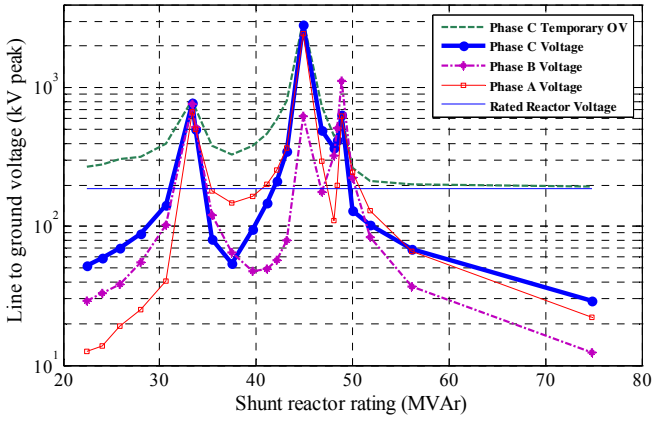


Figure 2

Three phase line to ground peak voltages and phase C temporary peak overvoltage versus three phase shunt reactor ratings: no-fault, the bottom line is opened and the other ones are energized

We put a generator at each terminal of the double-circuit line. These two generators have the nominal line voltages equal to 230 kV. About 60 MW active

power is flowing from right to left due to a small difference between two generators load angel.

The PSCAD has various transmission line solver models. In this case the analysis is for low frequency so the Bergeron model was picked as a solver which is a constant frequency model based on travelling waves [19], [20]. The results from the Bergeron model without damping approximation are with a good accordance with the ones from electrostatic matrix solution explained in the next section, so we used this option of Bergeron solver.

2.2 Analysis with Linear Reactors

Three values of shunt compensations may result in resonance on the opened untransposed line. These resonant conditions can occur in a no-fault condition and during the occurrence of any fault on the energized circuit as well. Faults in the energized circuit may lead to higher values of resonant voltage because of unbalancing, but they won't change the resonant points [13]. However, faults in the opened circuit can result in resonance for other shunt-reactor values. In previous works the faults on both the opened and the energized circuits were analyzed and up to 19 shunt-reactor values can cause resonance on an untransposed line [13].

To find these values, in the first step we changed the shunt linear reactors in a wide range and recorded the three phase line to ground voltages in no-fault condition. Both lines are energized in the first state and then the breakers in the bottom one in Figure 1 are opened and then the steady state line to ground voltages of each phase are recorded. A temporary transient overvoltage is observed after breakers operation anyway. Figure 2 shows the line to ground peak voltages values versus MVAr shunt reactor ratings in no-fault conditions with linear reactor simulation. Also, it contains the peak values of these temporary overvoltages recorded after breaker opening. This figure has a very good accordance with the previous works done by matrix solution [8], [11-14] and EMTP [16].

The corresponding resonant points are 33.3, 44.9 and 48.8 MVAr reactive compensations. The electrostatic analysis results are 30.8, 45 and 50.5 MVAr reactor compensations with the guard wire consideration. These values from the simulation are favorably comparable with the electrostatic no-loss solution. The details of the electrostatic method have been explained in the next section.

Faults on the opened line lead to different resonant points from no-fault ones [13]. Various faults comprise line to ground, line to line and line to line to ground was applied to the opened line and the line to ground peak voltages of the reactors were recorded. These voltages for phase C are shown in figure 3. There are two resonant points in LG fault on phase B: 38.7 and 44.9 MVAr reactive

compensations, two points in LL fault on phases A and B: 33.4 and 45.8 MVAr compensations, and one point in LLG fault on phases A and B: 42.1 MVAr compensation. The results from electrostatic solution are 42.7, 48.9, 31.2, 45.5, and 49.3 MVAr compensations respectively. Regarding to above outcomes, the simulation results are comparable with the ones derived from the electrostatic method.

This procedure should be done for two other situations. In the first one, the bottom line is opened and only one of the top lines is live. This situation leads to some resonant points as well. The diagram has not been reported here but in this situation the corresponding resonant points are 36.6, 46.7 and 48.1 MVAr compensations of the bottom line. In the second situation, the bottom line is live and the above lines are open. This yields 16.8, 22.5 and 24.4 MVAr shunt compensations as resonant points of the top lines. The corresponding inductor amounts of these reactors are approximately the same for the reactors which lead to a resonance condition in the bottom line. This was predictable, because the line length and shunt compensation in one of the top lines are half compared to the bottom one, so the same inductor amounts cause resonance. This procedure should also be done in fault conditions for choosing the right values of reactors regarding to the value of compensation which is appropriate for voltage controlling in terminals.

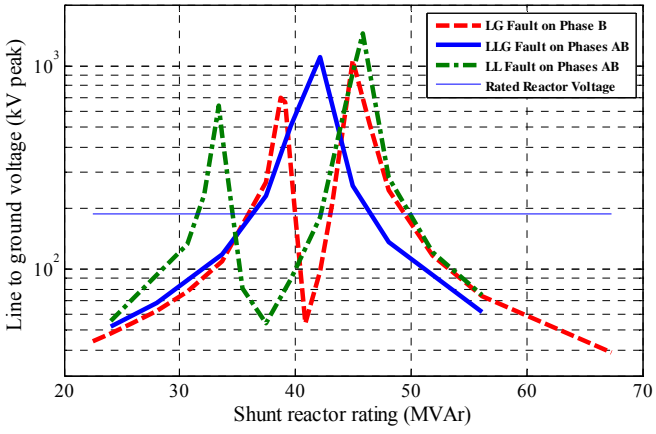


Figure 3

Phase C line to ground peak voltage versus three phase shunt reactor ratings with faults in opened line: LG fault on phase B, LLG and LL faults on phases A and B, the bottom line is opened and the other ones are energized

The safe range of shunt reactive compensation for bottom line according to the simulation results is below 27 MVAr and above 60 MVAr with15% certainty margin. This range is below 12 MVAr and above 32 MVAr for top lines separately with a 15% margin. The minimum and maximum resonant points for

bottom line from electrostatic analysis are 27.4 and 55.7 MVAr. With a 15% margin these values become 23 and 64 MVAr. Therefore, the electrostatic method gives a good approximation about the safe range of shunt compensation.

2.3 Analysis with Saturation and Corona Loss Considerations

The results derived from linear reactor simulation have a good estimation about the resonant points. However, it doesn't have a good estimation of resonance voltage levels. In some areas, the voltage levels reported in Figure 2 and Figure 3 are much higher than the reactor and line rating voltages. In this case two phenomena happen. The former is reactor saturation and the latter is corona dissipation due to the high level of voltages. Reactor saturation reduces the voltage over the reactor terminals because of changes in effective reactor value. But, it causes large currents going through the windings damaging the reactor as a result of excessive losses. As well, additional losses due to corona dissipations decrease the resulted voltage levels.

The PSCAD doesn't have a reactor with saturation capability. Therefore, for modeling a non-linear reactor we used a saturable unit transformer which has an inductor in secondary circuit. The parameters of transformer are set in a way that below nominal voltage the primary of transformer has the current equal to the linear reactor. The knee voltage of the reactor was selected equal to 1 pu, so above the nominal voltage the transformer becomes saturated and extra currents would go through the transformer. Because of air gaps in reactors the slope of B-H diagrams doesn't change very much in the saturation region. The flux versus current diagram of whole reactor and transformer is shown in the upside in Figure 4.

The corona has two major effects. First, the effective radius of the conductor increases resulting in changes in the capacitances of the transmission line [21], [22] and, second, the additional losses [21-24]. The former effect is important in fast transient analysis like lightning. However, it has an insignificant effect on steady state analysis like this work. But, the latter effect is important in our work.

The routine way of modeling corona losses is connecting a shunt resistor to the modeled circuit. There are two major ways of this resistor determination. One of them is for fast transients and is determined from the real time voltage of conductors and has different coefficients for positive and negative waves [23]. The other one is for steady state analysis and declares the resistor from the RMS voltage of conductors for modeling the average losses of corona [24]. We used the second one in the simulation. We connected six shunt resistors to the lines and the values of them are set real-time from RMS line to ground voltages. We determined the resistors in a way that the corona mean losses for 1 km of a conductor equal to the formula given in (1) [24].

$$p = 241(f + 25)\sqrt{\frac{r + 6/s + 0.04}{s}}(e - e_0)^2 10^{-5} \quad (1)$$

p = the loss per kilometer of conductor in kW
f = the frequency
e = line to ground RMS voltage in kV
e_0 = disruptive critical line to ground voltage in kV
r = the radius of the conductor in cm
s = the distance between conductor centers in cm

Figure 4 shows the resulted resonance voltage of phase C in no-fault condition versus the three phase shunt reactor ratings from simulations with linear and non-linear reactors. The voltage levels have dropped very much due to the saturation and corona losses.

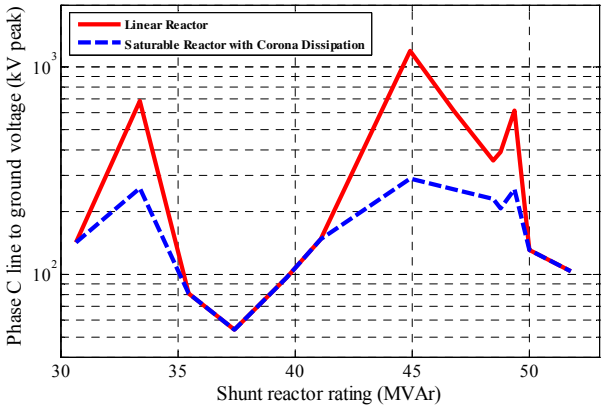


Figure 4

The opened line phase C line to ground peak voltage versus three phase shunt reactor ratings: linear reactor and non-linear with corona loss simulation results, the bottom line is opened and the other ones are energized. The upside diagram is Flux versus primary current of the saturable transformer.

2.4 Electrostatic No-Loss Analysis

In multi-conductor system, a capacitance matrix can be defined. First, the potential coefficient matrix (P) should be declared. It can be easily calculated based on the radiuses of conductors and the spaces between them [8, 25, and 26].

$$V = PQ \quad (2)$$

In this stage, the effect of ground can be participated with electromagnetic mirror rule. By reversing the potential coefficient matrix the capacitance matrix is derived.

$$Q = CV \;, C = P^{-1} \tag{3}$$

In a n-conductor system, the capacitance matrix would be in the order of n. If the order reduces to six, which is the number of line conductors (without guard wire), the capacitance matrix can be separated into four parts, as in (4) and after that the potential of the opened line conductors can be derived from the potential matrix of energized line conductors, regardless of the currents [8, 12, 27, and 28].

$$\begin{bmatrix} i_I \\ i_{II} \end{bmatrix} = j\omega \begin{bmatrix} C_{I-I} & C_{I-II} \\ C_{II-I} & C_{II-II} \end{bmatrix} \begin{bmatrix} e_I \\ e_{II} \end{bmatrix}$$

$$\left[e_{II}\right] = -\left[C_{II-II} + \frac{1}{j\omega} Y_{II-II} \right]^{-1} \left[C_{II-I}\right]\left[e_I\right] \tag{4}$$

i_I = current matrix of the energized line
i_{II} = current matrix of the opened line
e_I = potential matrix of the energized line
e_{II} = potential matrix of the opened line
$Y_{II\text{-}II}$ = the admittance matrix of the reactors connected to the opened line
ω = 2*pi*frequency

Without loss consideration, the opened line potentials go to infinity when the resonance occurs and it happens when the determinant of the first matrix becomes zero. So the equation that gives the resonant points is:

$$\det\left(\left[C_{II-II} + \frac{1}{j\omega} Y_{II-II} \right] \right) = 0 \tag{5}$$

In a six-conductor system like a double-circuit transmission line without a guard wire, the capacitance matrix can be separated into four 3*3 matrixes and then the resonance points can be calculated. In a double circuit transmission line with guard wires, the order of capacitance matrix is above 6 and for using this method we should decrease the matrix to the order of 6. We can use the principles of the capacitance matrix to reduce the order of the matrix.

In a capacitance matrix, the non-diagonal elements are the minus of the capacitors between corresponding conductors (6). The diagonal elements are the sum of the capacitors related to a conductor including the capacitance of that matrix to ground (7).

$$C_{ij} = -c_{ij} ; i \neq j \tag{6}$$

C_{ij} = non-diagonal element of the capacitance matrix

c_{ij} = capacitor between conductor i and j.

$$C_{ii} = c_{ig} + \sum_{i \neq j} c_{ij} \tag{7}$$

C_{ii} = diagonal element of the capacitance matrix
c_{ig} = capacitor between conductor i and ground
c_{ij} = capacitor between conductor i and j.

So in a multi-conductor system with given capacitors, the capacitance matrix can be calculated using (6) and (7). In a system with a guard wire, regardless of the guard resistor, the wire has the potential of ground. This approximation is good especially in this case which we deal with low frequencies. So like the six-conductor system, we have seven potential references: six conductors of double-circuit and ground. For decreasing the order of the matrix, we should add the capacitors between each conductor and guard wire to the corresponding diagonal element because this capacitor is between a conductor and a guard wire having ground potential. After decreasing the order of the matrix, it can be separated and resonance points can be derived.

This method also can be used in fault conditions [13]. For example, when we have a line to ground fault on a conductor of the opened line, the potential of that conductor would be ground and we would have six potential points besides seven ones. We can add the capacitors between other conductors and the faulty one to the diagonal elements and then the matrix can be separated into four parts. In this case we have two unknown potentials on the opened line, so the $C_{II\text{-}II}$ would be 2*2 and therefore we have two resonant points in this case.

When we have a line to line fault, we have again six potential points because the potentials of two conductors, which are faulty, would be the same. The equivalent element of these two conductors in the capacitance matrix would be the sum of diagonal corresponding elements minus two times of the capacitor between them [13]. In this case, the admittance matrix would be different because one phase of reactor is on the conductor of the opened line which is not faulty and two parallel phases of the reactor are on the two conductors having a fault.

As a numerical example, the $C_{II\text{-}II}$ matrix is given in (8) when the bottom line is opened and the top ones are live. By solving the (5) for this matrix, the resonant points will be 5.4, 3.7 and 3.3 H corresponding to 30.8, 45 and 50.5 MVAr in 230 kV. The other resonance points for fault conditions can be calculated respectively by this method.

$$C_{II-II} = \begin{bmatrix} 2.451 & -0.401 & -0.173 \\ -0.401 & 2.512 & -0.417 \\ -0.173 & -0.417 & 2.642 \end{bmatrix} (\mu f) \tag{8}$$

3 Methods of Damping the Resonance Condition

Several possible routine avenues are possible for correcting the resonance situation. Assuming the reactor rating, reactive capability, stability, and light load or open circuit voltage limitations, the following alternatives are present [8]:

1) Complete transposition of the circuits.

2) Ungrounding the neutral of the reactor.

3) Insertion of a resistor between ground and neutral of the reactor.

4) Insertion of a reactor (either directly connected or coupled via a transformer) between ground and neutral of the reactor.

The solution works if it decreases the resonance voltages over reactor terminals in no-fault and faulty situations in three configurations: the bottom line is open and the top ones are live, the bottom line is open and one of the top lines is live, the bottom line is live and one or both of the top lines are open.

The first alternative may not be economically justified, especially when the circuit is constructed like here. The last three alternatives require analysis not only to ascertain the effectiveness of reducing the line-to-ground voltages, but since reactors have graded insulation, to determine that the neutral-to-ground voltage on the reactor is within the insulation specifications.

Ungrounding the neutral of reactor does not have any considerable effect. Moreover, because of low neutral current in resonance situation, adding a resistor or a reactor in the neutral of the reactors does not help much and the reactor terminal voltages are still higher than the rated ones. Therefore, none of these routine solutions work in present circuit and new ways should be found to reduce the resonance voltages.

A possible way that is routine but has considerable cost is to add three 25 MVAr reactors in the system; one in Zahedan-Iranshahr line in Zahedan substation and the others in Khash substation, on Khash-Iranshahr and Zahedan-Khash lines as shown in Figure 5 (a). In this situation the bottom line will have overall 75 MVAr reactive compensation and each top line will have a 50 MVAr compensation; so the circuit will go far from resonance situation. In addition, when two lines are live, the compensation in Zahedan and Iranshahr substation will be 50 and 75 MVAr respectively and the voltage does not reduce too much because of high shunt compensation. The reason of choosing 25 MVAr reactors is that this type of reactors is routine and available in the Zahedan power section. Lower reactor ratings like 10 MVAr reactors can be added instead of 25 MVAr reactors, but this needs to build and design new reactors and so that buying three 35 MVAr reactors and replacing with the present ones will have the better reliability for system and the 25 MVAr reactors can be used in other places, so it is more beneficial.

Another way which has lower cost is replacing the ground disconnector switches with breakers. In this way, when one line is opened, the breakers can be closed and earth the system. Earthing with disconnector switches can be dangerous because high voltage induction on the line can cause arcing between switch terminals. This solution has a disadvantage. In the time between opening the line breakers and closing the earth breakers the reactors will withstand the temporary over voltages and this can damage them, but if this time decreases this overvoltage will be eliminated and the reactors will not be stressed out. Another disadvantage is that in this method the reactor windings suddenly become short circuited. If the windings have major voltages before the breaker closure it can damage the winding insulation due to the high voltage variation and non uniform voltage distribution over winding loops. Also the current that goes through the circuit breaker has a decay DC with large time constant and goes to zero in about 20 seconds, so the breaker must not open before this time because the current has no zero crossing. Adding a resistor in series with breaker can lower the time constant and the voltage variation stress on reactor windings.

Another way, which doesn't need any new equipment, is to change the reactor configuration in a way that no resonance happens. Figure 5 (b) shows the new proposed configuration. In this configuration, the Zahedan-Iranshahr line has 75 MVAr shunt compensation and has no strict resonance in open condition. The top lines have no shunt compensation in dead condition and have no resonance as well. Moreover, when the top line is connected, it has a 25 MVAr reactive compensation and the reactor is located in the middle of the line, so the overvoltage of half of the top line is not so high. Substations have also other lines with reactive compensation and there isn't a severe overvoltage in the light load situation. Figure 6 shows the voltage on the bottom line when it opens at 1.5 s and the top lines are connected.

Another strategy is using a TCR configuration with resistor instead of an indicator for damping the resonance condition. The proposed layout is shown in Figure 7 (a). When the line to ground voltage of a line becomes higher than 1.2 pu, then a controller can increase the duty cycle of switching and intensify the effective resistor resulted in the decrease of resonance voltage. A wide range of resistors will lower the resonance voltage below the rated terminal level. As an example in the presented circuit, a 270 Ω resistor can reduce the resonance induction below the nominal voltage of rector terminals and the corresponding power loss is about120 kW. By using the substation internal feeding transformer type as the TCR transformer, it doesn't need to design and buy a special transformer and the cost will decrease. These types of transformers have the rated 300 kVA nominal power so are suitable for this application. These transformers have the ratio of 230 kV/400 V, so there isn't any need to high voltage thyristors and the high current low voltage ones, which are very usual, can fit into this structure. The main disadvantage of this solution and the ones which need adding equipments are lowering the system reliability.

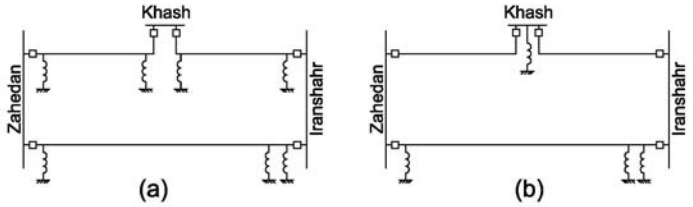


Figure 5
The new configuration proposed for lowering resonance voltage levels: (a) by adding new reactors, (b) using the available reactors

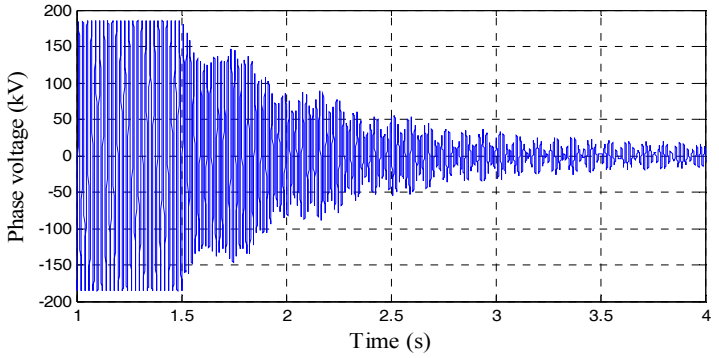


Figure 6
Damping of resonance in phase C with re-configuration

Another possible approach is to insert a transformer like the previous solution but connecting a capacitor bank in the secondary of the transformer. In this way the voltage level will decrease very much, if a suitable capacitor is chosen. The considerations in this solution are the capacitor bank current, switching the transformer and ferroresonance of the transformer and the capacitor. The voltage reduces due to equivalent capacitor seen by the system but there is a current which is related to residual voltage on the system. If the capacitor selection will be correct the residual voltage will be very low and the current of capacitor bank will be limited. In this case, insertion of a 3 MVAr capacitor bank in the secondary of a 230 kV/ 20 kV transformer just in one side of the bottom line will decrease the all reactor terminal line to ground voltages to 15 kV peak from 240 kV peak in the absence of the capacitor bank and the steady state current of capacitor bank will be 110 A RMS in the worst phase. The switching of capacitor bank always has some transients, but in this situation switching is done through the transformer and the equivalent inductor of transformer limits the transient overcurrent. In the presented case, using a regular transformer with 0.1 pu series impedance, the first peak during switching becomes 55 A when the steady state current of the primary of the transformer is 10 A RMS. The last consideration is ferroresonance of the

transformer with the capacitor bank. When the transformer is connected to the system, the ferroresonance happens but the voltage peaks aren't too high and by adding a resistor in series with the capacitor banks, this ferroresonance will damp. Therefore, there is no strict stress over transformer and the capacitor bank. In this case, a 1 Ω resistor in each phase will damp the ferroresonance in about 5 seconds and the power loss of each of them during connection to the circuit is about 12 kW. In this procedure, the reactor will be stressed out during the time between the opening of the line breaker and the connection of the capacitor bank. Figure 7 (b) shows the detail of the configuration of this solution. Figure 8 shows the voltage on the bottom line when this solution has been applied. In this figure the top lines are energized and the bottom line opens at 1.5 s, then, the capacitor bank connects at 5 s. It is obvious that there is a slight ferroresonance but it damps within 2 s and the steady state voltage of line is lower than the re-configuration method.

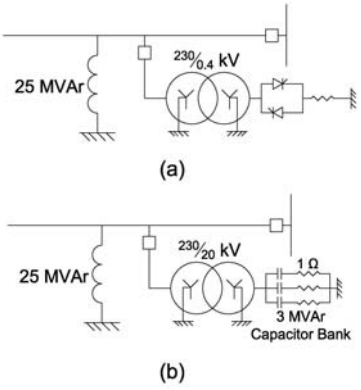


Figure 7

Proposed configurations for damping the resonance circuit: (a) using a switching resistor, (b) using a capacitor bank

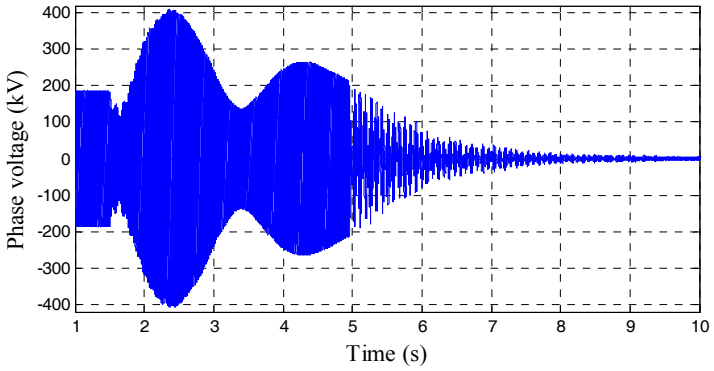


Figure 8

Damping of resonance in phase C with capacitor bank

Among all above approaches, changing the configuration of reactors (Figure 5 (b)) is the best solution, because it doesn't need any additional equipment and it is cheaper than others. Also the reliability of the final system will be the best compared to the other solutions.

Conclusions

1. Choosing the reactors rating in a double circuit transmission line needs special considerations of parallel resonance probability.
2. For finding the resonance points, all available configurations of the line including various faults on the energized and opened line should be considered.
3. Faults on the opened line leads to different resonant points, but faults on the live circuit changes the resonance voltage levels.
4. The electrostatic no-loss solution gives good estimation about the resonance points and can be used for selecting the appropriate shunt reactive compensation.
5. Modeling the system in PSCAD using Bergeron transmission line model without damping approximation leads to answers which are favorably comparable to the ones from electrostatic method and can be used for reactors rating selection. As well, Reactor saturation and corona losses can be modeled to get better results.
6. In resonance situations which the neutral of reactor current is low, insertion of reactor or resistor in the neutral of the reactors does not have considerable effect on the voltage levels.
7. The TCR configuration with appropriate resistors can be used to lower the resonance voltage.
8. Insertion of suitable capacitor can highly affect the resonance condition and can be used for damping the resonance circuit, but needs special considerations.
9. In resonance condition the opened line can be earthed by using circuit breakers instead of the earth disconnector switches.

References

[1] G. XinBo, W. YingShi, Q. Tong, X. Wei, L. Daoning: The Simulation of the Controllable Reactor and It's Application in Ultra High Voltage Transmission Lines, Int. Conf. Advanced Power System Automation and Protection, TBEA Shenyang, China, 2011, Vol. 3, pp. 1833-1837

[2] R. Begamudre: Extra High Voltage AC Transmission Engineering, New Age International, 2007

[3] Z. Li, X. Yuqing: Application and Development of Shunt Reactors in EHV & UHV Transmission Lines, Electric Power Automation Equipment, 2007, vol. 21(4), pp. 18-24

[4] X. Qiufeng, W. Haiyan, W. Zhiwei: Application and Development of Shunt Reactors in EHV & UHV Transmission Lines,Guangdong Power Transmission Technology, 2007, Vol. 24(4), pp. 4-6

[5] P. Zhendong, Z. Jiawen: Power Frequency Over-Voltage of 500 kV Four-Circuit Lines on the Same Power, East china electric power, 2007, Vol. 35(3), pp. 24-27

[6] G. Dingxie, Z. Peihong: Over-voltage, Secondary Arc and Reactive Power Compensation in UHV AC Transmission System, High Voltage Engineering, 2005

[7] C. Hansheng, H. Danhui, T. Caiqi: Study on the Shunt Comensation and Overvoltage in Zhengnan 500 kV System, High voltage engineering, 2000, Vol. 26(5), pp. 34-37

[8] M. Hesse, D. Wilson: Near Resonant Coupling on EHV Circuits: II - Methods of Analysis, IEEE Trans. Power Apparatus and Systems, vol. PAS-87, Feb. 1968, No. 2, pp. 326-334

[9] X. Lv, Q. Sun, Q. Li, W. Shi: Multi-objective Parameter Optimization of Shunt Reactors for Multi-circuit Transmission Lines on the Same Tower, in Proc. 4th Int. Conf. Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), 2011, pp. 271-276

[10] M. Pickett, H. Manning, H. Van Geem: Near Resonant Coupling on EHV Circuits: I – Field Investigations, IEEE Trans. Power Apparatus and Systems, Vol. PAS-87, Feb. 1968, No. 2, pp. 322-325

[11] A. Chaston: EHV AC Parallel Transmission Line Calculations with Application to the Near Resonance Problem, IEEE Trans. Power Apparatus and Systems, May. 1969, Vol. PAS-88, No. 5, pp. 627-635

[12] K. Priest, A. Ramirez, H. Howak, J. Laforest: Resonant Voltages on Reactor Compensated Extra-High-Voltage Lines, IEEE Trans. Power Apparatus and Systems, Nov. 1972, Vol. PAS-91, No. 6, pp. 2528-2536

[13] E. E. Colapret, W. E. Reid.: Effects of Faults and Shunt Reactor Parameters on Parallel Resonance, IEEE Trans. Power Apparatus and Systems, Feb. 1981, Vol. PAS-100, No. 2, pp. 572-584

[14] W. E. Reid, R. F. Gustin, P. V. Zylstra: Guidelines for Determining Parallel Resonance on EHV Transmission Lines, IEEE Trans. Power Apparatus and Systems, Sep. 1983, Vol. PAS-102, No. 9, pp. 3196-3204

[15] M. H. Hesse, J. Sabath: EHV Double-Circuit Untransposed Transmission Line-Analysis and Tests, IEEE Trans. Power Apparatus and Systems, May 1971, Vol. PAS-90, No. 3, pp. 984-992

[16] M. V. Escudero, M. Redfern: Parametric Analysis of Parallel Resonance on Shunt Compensated Transmission Lines, in Proc. 39th Int. Conf. Universities Power Engineering, UPEC, 2004, Vol. 2, pp. 1181-1185

[17] L. Wei, N. Wen-hui, H. Dong-shan: Analysis and Modification of a 500kV Transmission Line Overvoltage Problem, in Proc. 2010 China Int. Conf. Electricity Distribution (CICED), pp. 1-6

[18] A. B. Fernandes, W. L. A. E. Neves, G. Costa, M. N. Cavalcanti: The Effect of the Shunt Conductance on Transmission Line Models, in Proc. Int. Conf. Power System Transients, Rio de Janeiro, Brazil, 2001, pp. 49-54

[19] PSCAD Online Help, v4.2.1, Manitoba HVDC Research Centre Inc., 2006

[20] H. W. Dommel: Digital Computer Solution of Electromagnetic Transients in Single and Multiphase Networks, IEEE Trans. Power Apparatus and Systems, Vol. PAS-88, No. 4, pp. 388-399, Apr. 1969

[21] A. R. Hileman: Insulation Coordination for Power Systems, Boca Raton: CRC-Taylor & Fransis Group, 1999, ch. 9

[22] J. A. Martinez-Velasco: Power System Transients Parameter Determination, Boca Raton: CRC-Taylor & Fransis Group, 2010, ch. 2

[23] J. C. Das: Transients in Electrical Systems Analysis, Recognition, and Mitigation, McGraw-Hill, 2010, ch. 4

[24] F. W. Peek: Dielectric Phenomena in High Voltage Engineering, McGraw-Hill, 1915, ch. 5

[25] D. K. Cheng: Field and Wave Electromagnetics, Addison-Wesley, 2nd edition, 1989, ch. 3

[26] H. Saadat: Power System Analysis, McGraw-Hill, 1999, ch. 4

[27] E. T. B. Gross: Unbalances of Untransposed Overhead Lines, J. Franklin Inst., 1952, Vol. 254, pp. 487-497

[28] E. T. B. Gross, M. H. Hesse: Electromagnetic Unbalance of Untransposed Transmission Lines, Trans. AIEE (Power Apparatus and Systems), Dec. 1953, Vol. 72, pp. 1323-1336

A Framework for Delivering e-Government Support

Goran Šimić, Zoran Jeremić

Military Academy, University of Defense
Pavla Jurišića Šturma 33, 11000 Belgrade, Serbia
e-mail: goran.simic@va.mod.gov.rs; zoran.jeremic@va.mod.gov.rs

Ejub Kajan

State University of Novi Pazar
Vuka Karadžića bb, 36300 Novi Pazar, Serbia
e-mail: kajane@acm.org

Dragan Randjelović

Academy of Criminology and Police Studies, Dušanova 196 Street, 11080 Belgrade, Serbia, e-mail: dragan.randjelovic@kpa.edu.rs

Aaron Presnall

Jefferson Institute, Kneginje Ljubice 28, 11000 Begrade, Serbia
e-mail: apresnall@jeffersoninst.org

Abstract: This paper gives a solution for improving e-Government services based on a hybrid approach: multilayered clustering of e-government documents based on fuzzy concepts and application of different text similarity measures. The goal is to reduce time between citizen's questions and government feedback, either completely eliminating or at least minimizing the deployment of subject matter experts. After the problem description, the paper describes step by step the functionality of the proposed system. At the end, concluding remarks emphasize some important features of the given approach and potential for future research.

Keywords: e-government; clustering; text similarity; fuzzy clustering

1 Introduction

Citizens' access and right to information at the level of local government is one of the essential ingredients for a successful government. Access to information empowers citizens to make decisions on the issues of government that affect them, decisions which provide critical feedback toGovernment as it seeks to meet the needs of citizens and improve their quality of life. Government should actively seek to capture the positive feedback loop inherent in providing greater access to information as a critical component of its strategy to deliverhigh qualityjust in time services for citizens and businesses. The use of advanced information technology to provide easier access to public information and government services is therefore a necessary condition for good governance. However, a number of challenges must be addressed to fully utilize the benefits of available technology.

A growing volume of information related to government rules, regulations, amended provisions, legal precedence and interpretive guidelines are distributed on a multitude of government portals, so that citizens can browse, search and take action. Some of these portals are equipped with search engines that provide text based search of documents. However, government documents are often very long andwithmany cross references to other related documents. Moreover, these documents are semi-structured with similar and often ambiguous content and terminology when taken as isolated texts out of context. As such, the characteristics of government document records make simple text search a serious impediment to understanding and use by common citizens.

Moreover, most of these portals are based on a one-way relation, in which the government produces and delivers information for use by citizens. This information is categorized or could be searched through a simple search engine providing keyword based search. The results of such a search could be a large number of documents the citizen has to go through to find desired information. If his knowledge in law and policy is limited, it could take hours to find the appropriate information.

Despite considerable attention to the introduction of ICT in government, most developed and developing countries have so far focused on the relatively easy phase of e-Government: developing websites, piloting a few applications, and putting these services online. Developed countries have been better able to invest in ICT infrastructure and service improvement, while developing countries must carefully evaluate the marginal utility of such investment. While the global trend remains one of steady improvement of e-government services in all regions, there is a growing gap of e-government development between developed and developing countries [2].

The 2012 world leader in e-government development was the Republic of Korea, which uses a single government portal as a gateway to services from multiple channels, organized by theme and subjects [2]. Many departments are integrated

together through a powerful search engine offering an advanced categorizing function, which can list results by websites, services, and news. Mexico takes a different approach to integrate services provided to citizens. It provides a search engine that respond to users' specific search criteria, which has ability to filter information based on the information type, theme or user's location. Serbia significantly increased the performance of its e-government recently. The Digital Agenda Authority is responsible for introducing online services to improve the quality of services provided to citizens based on the "all services from one place" principle. The Authority created a portal, eUprava (http://www.euprava.gov.rs), which aggregates services and information from more than 27 governmental authorities, including municipal authorities.

Most countries from the European Union follow the approach of separate portals for their information, service and participation offerings. However, a recent trend in many countries is to set up portals that aggregate large amounts of information and services into a single website. A common approach includes organizing content around themes and/or specific audiences. These portals include search features that may index content from other government websites.

This paper proposes a novel approach to facilitate and foster e-government optimization and automation through the use of advanced information retrieval methods and techniques. In the next section, we describe the problem of the existing e-government solutions in Serbia. Section 3 describes a proposed alternative solution and a use case. Afterward, we provide a description of the design and implementation of the proposed solution. We conclude this work with a description of the potential benefits of the proposed alternative approach.

2 Problem Description

Serbia has achieved great improvement in the development of e-government over the last few years. At the beginning of 2007 a central portal of e-government services in Serbia was created (www.euprava.gov.rs). The main goal of the portal development was to provide a common access point for all e-government services provided to citizens, companies and public administration. Through this portal, citizens can download all relevant service documents, find links to a full range of public institution Web sites, and learn more about how to use e-government services in general.

However services for citizens and the private-sector in Serbia have not yet gone beyond a nominal level. Citizens can download forms; get laws, reports and other related publications. The information flow is unidirectional, from government to citizen, and if there is a need for a bidirectional flow of information, most institutions encourage communication by e-mail, where response times vary widely. Even though e-government has come a long way, there is room for

improvement especially in the quality of provided services, including better knowledge management and improved two-way interaction between government and citizens.
There are two objectives that should be addressed when assessing e-government system in Serbia:

- E-government services should be designed around users' needs and provide advanced search capabilities that will enable better and easier access to information;
- E-government should enhance government services by reducing the administrative burden, improving organizational processes and using ICT to improve efficiency in public administration.

Search features provided by an e-government portal are commendable, but do not meet either of these objectives. Citizens without expert knowledge in the domain of inquiry are often disappointed by the difficulties that must be overcome and efforts they have to make in order to access or gather the requested information, and ultimately by the lack of effectiveness in orchestration of the various procedures. Domain experts must be engaged to examine the various cases and select the appropriate service or information requested by citizen. Typically, such a scenario would consist of the following steps: a citizen makes a request for specific information elaborating her specific case through the use of email or phone; a government officer receives the request, examines the request, clustering the nature of the problem and sends it to a specific domain expert; the domain expert evaluates the citizen's case and prepares the requested information.

Some of the citizen's cases are obvious, related to previous cases or they are clearly and fully described in one document. However, collecting and consolidating all the information could be very difficult and time consuming. The domain expert may have to search across many documents, to search for relations between documents and the specific case or to compare the case with previous solved cases.

Moreover, the limited number of government employees limits the number of citizen requests that could be processed without significantly increasing costs. In addition, most of the services that the government portal offers declare a wait time of 2-6 weeks, which is not acceptable. The transformation to the improved, more intelligent system could deliver a drastic reduction on the average response time for a request from a citizen.

E-government should provide a solution to manage a citizen's requests by documenting and tracking it through to the final resolution. It should use previous cases and resolutions to provide a faster, smarter and more efficient response to a citizen's request. Only in specific and new cases or if the citizen is not satisfied with the proposed resolution, should the human government officer be asked to intervene.

The approach we suggest aims to enable interactive processes that are simple, effective, and based on the user's needs and capabilities, rather than the government's organizational structure or government business models. It should create the opportunity to evaluate and eliminate redundant or unnecessary steps and processes as well as to reduce costs and cycle times by transitioning from the processes mainly based on human-related work to automated and more intelligent user centered processes.

3 Advanced Answering Engine (ADVANSE) for Interactive E-Government Services

E-government systems under consideration here are designed to assist citizens in making decisions. Citizens ask questions and the system tries to response with an appropriate answer to inform the citizens' decision or next step. As described in Section 1, in the current e-government systems of Serbia, these questions are answered by Subject Matter Experts (SMEs). The response time varies from one to several days, depending on the availability of the SME. On the other hand, a number of questions and answers accumulate over time. They could be considered as a kind of the knowledge base (KB). This KB could be captured and applied as a good basis for development of the advanced answering engine (ADVANSE).

3.1 Scenario of Use

Interaction between the citizen and the system happens in three phases (Figure 1). At the beginning, the system offers groups of key terms (phrases and words) to the citizen. The citizen can select one (or not select any) according to her question. In the next step, the system delivers her the set of questions that are strongly related to the selected key terms. The citizen has two options in the second phase: to choose a question from the list, or to enter a new one. Regardless of which option is chosen, the system delivers an answer in the next phase.

If the citizen selected the 1st option, the system delivers an answer that is fully matched to the selected question. Otherwise, the system finds an existing question that is the most similar to the new one. The system sends this question's answer back to the citizen. The citizen can evaluate this answer. After this three steps interaction, she can continue, or finish the session with the system. If she is not satisfied with the answer she can try to find another question or enter the new one.

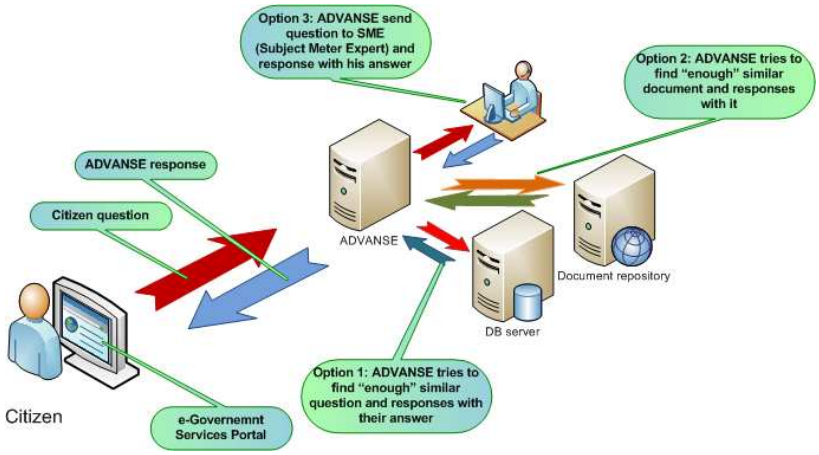


Figure 1
Interaction between citizens and system

The case in which the citizen writes a new question represents the focus of the research because the system tries to find the most appropriate answer. The groups of key terms (mentioned in the first step of interaction) are generated automatically by a clustering process. The questions and formal documents belong to these clusters (according to similarity of the terms and words they contains). When the citizen enters a new question, the system calculates the similarity with questions in the determined cluster. If the similarity threshold is satisfied, the system delivers the best fitted answer to the citizen. Otherwise, it can deliver the related document and / or the safety answer (usually it is a message with an appropriate explanation, recommendation, or references to other resources). The citizen can follow the steps offered, or change the way of interaction.

3.2 Content Representation

There are three basic concepts (content) in the system: citizen questions, governmental documents, and answers. They are separated into different layers (Figure 2). The answers' layer is between questions and documents (QD) layers. QD are clustered by key terms. There are as many clusters as key terms in the domain dictionary. The answers are not clustered because of two reasons: they are excluded from searching and their concept has double purpose. An answer can be manually created by a subject matter expert or automatically generated by the system by making an association between the question and related documents. Formed associations can be of different types. Each question can be related to one or multiple answers and each answer can be relevant for one or more questions. The same approach is applied for associations between answers and documents.

Questions and documents can contain more than one key term. If the content is large or more general, there is more chance for it to score a hit. The question or

document then belongs to more than one cluster. Therefore, the clusters can be represented as the sets that are intersected with each other and the questions and documents that contain more than one key term belong to these intersections.

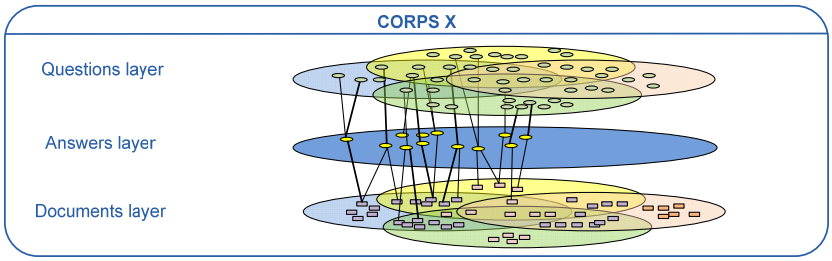
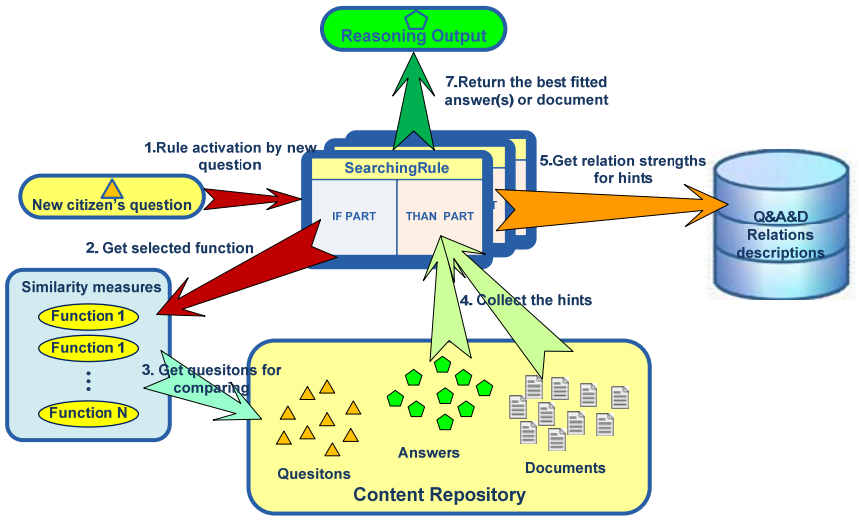


Figure 2
Layered content representation

The strength of relations between the content from different layers depends on citizens' satisfaction with the feedback. This value is calculated from two factors. The first one is degree of similarity between nodes and another is from the citizen evaluation of the system response. The degree of similarity between the answers and documents is calculated by the system and it represents an objective value (e.g. cosine similarity). The strength between questions and answers are continually changing and depending on the score given by citizens. This is a kind of content evaluation which depends on citizens' individual expectations and attitudes. Therefore, this measure has a subjective nature, but it becomes more objective with more citizen feedback.

3.3 The Role of the Rules

Besides the questions, answers and documents, rules represent another part of the system knowledge (Figure 3). Rule based reasoning is used for two purposes: separating business logic from heterogeneous data and separating the similarity measuring from decision making. There are two rule types in the system: searching and creating rules. Searching rules provide a flexible way for reasoning on similarity between questions and existing content. The final decision about responding to an answer or document is the result of rule based searching. Searching rules are designed for finding questions and answers existing in the system similar to the citizen question. There can be more than one similarity measure (algorithm) in the system. They are implemented as functions which are invoked by statements in the rule premise. This way the reasoning about similarity can be changed by using different algorithms. Further, the function compares the new question with the existing ones and returns the best fitted question. This question is forwarded to the rule action part as a parameter of another function there. This function finds and returns the appropriate content (answers or documents) that represents the final result of the reasoning.



Fi

Figure 3

Relations between rules and system concepts

The creating rules are designed for making relations between questions, answers and documents. Both types have the question sets on the left hand side and sets of triplets (questions, answers and their relations) on the right hand side. The rules are generated by the system by using those data. In contrast, creating rules are used when a new relation between question and answer has to be established. These rules do not change the content but the connections between questions and answers (see the last paragraph in Section 3.2).

The system complexity is reduced by using the rules. The complex functions designed for different purposes (measuring similarity, for calculating the relation strength and clustering) are embedded. The rules just contain the function calls in the premises, or in the action parts.

4 Design and Implementation

The main focus of the conceptual model is the citizen's question (Figure 4). This question can be in relation with one or more answers, or/and one or more documents. The question is presented in the model with the Citizen Question concept, whether it has been answered or not. This is a main concept in the system because processing of questions is the top objective of the system.

The questions and documents belong to the clusters. The answered questions are in associative relations with the answers. The answer concept has a dual nature. It can consist of the text written by an SME or automatically generated

recommendations – links to the documents appropriate to the citizens' question. In this case it depends on document(s). The relations between answers and questions are weighted. The strength of the relation has a default (initial) value that is changeable by the citizens' feedback about her satisfaction with the answer. Therefore, relations are represented by the QASTriplet (question – answer – relation strength) concept.

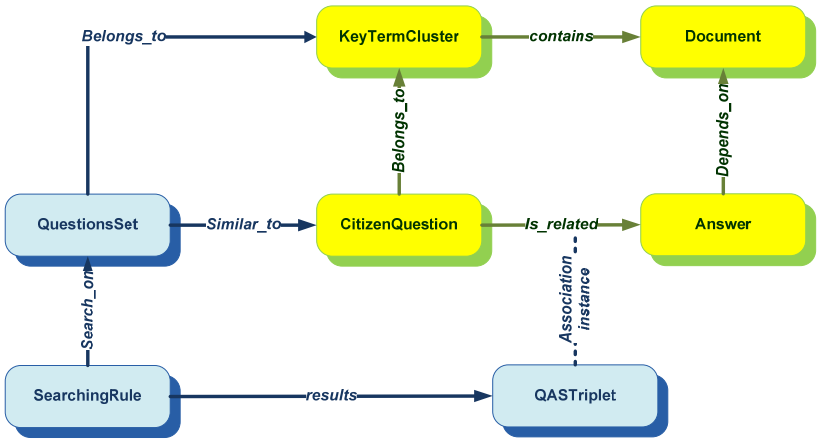


Figure 4
Basic conceptual model

Another part of conceptual model is designed for reasoning purposes (blue colored). The central concept in this section is the Searching Rule. As mentioned in the previous chapter, rules are used for providing flexibility to reasoning. The set of questions, which is searched by the function embedded in the rule's left hand side, is represented by the Questions Set concept. Since different similarity functions (measurements) can be used, the search results can be different and overall system behavior is flexible for that reason. The purpose of finding the most similar question to the new one is to get a related question – answer – relation strength triplet. This data structure is represented by the QASTriplet concept. If the rule is fired, the function on the rule's right hand side returns the triplets that are related to the resulted question. Detailed discussion about how the system uses the rules is in the following section (Section 4.3).

4.1 Clusters Generation

The clustering initialization could be performed in different ways: Random partition [3], *Forgy* partition [1] and *Kaufman* [6] are the most commonly mentioned. The *Kaufman* method does not need a predefined number of clusters, but the other two do. In ADVANCE the number of clusters is predefined: it is determined by the number of used key terms. Questions and documents are fuzzy clustered. The developed method (Equation 1) is based on the ideas of *fuzzy c* –

means (FCM) algorithm [5], [13]. Different of FCM, a concept of distances is avoided because key terms are used as constant values instead of the iterative calculation of some statistical value of central tendency every time the set of observations is changed (Equation 1).

$$f_{fcm} = \sum_{i=1}^{N} \sum_{j=1}^{K} m_{ij}(x_i) \tag{1}$$

In this way, a membership function (m_{ij}) of document or question (x_i) represents the only measure of its belonging to the particular (j-th) cluster. It is calculated just one time and there is no need for recalculation every time a new question or document is added into the system. This approach is found useful because there is a need for permanent adding of new questions into the system.

The multiplicative nature of the algorithm expresses the fact that every data portion belongs to every cluster in some degree. If there are K clusters and N questions or documents (they belong to separate layers), the $K \times N$ matrix of values of the membership function is formed. Considered dynamically, the addition of a question or document produces the $K \times (N + 1)$ matrix – one column is added into the existing matrix.

ADVANSE calculates the membership function $m_{t,q}$ (where t represents the clusters' key term and q represents the question or document) by using both the term frequency ($tf_{t,q}$) and the inverse document frequency ($idf_{t,q}$) [11], [8] (Equation 2). The first ($tf_{t,q}$) represents the internal characteristic – how many occurrences ($f_{t,q}$) of the specified term t there are in the particular content q (question or document). The other measure ($idf_{t,c}$) is on the global level – how many questions that contain the specified term ($N_{q,t,c}$) are there in the whole corps ($N_{q,c}$). The product of these two values is commonly called the TF-IDF function.

$$m_{t,q} = k \cdot tf_{t,q} \cdot idf_{t,c} = k \cdot \log(f_{t,q} + 1) \cdot \log \frac{N_{q,c}}{N_{q,t,c}} \tag{2}$$

Logarithm functions are used for normalization purposes. This way the membership function value varies in range from zero to one. There is an additional coefficient k – correction factor that provides better dispersion of the values of $m_{t,q}$ in the range. This coefficient is calculated as the reciprocal of the TF-IDF function maximum.

The described method provide for flexible behavior of the system. Every new question or document added into the system can be processed particularly. There is no need for repeating the clustering initialization completely. The membership values are calculated and simply stored as metadata ready for filtering purposes (e.g. a question or document can be taken under consideration depending on the

threshold – changeable minimum value of the membership function). Only changes in key terms will produce re-initialization of clusters.

4.2 Algorithm Description

The proposed algorithm is described with an activity diagram (Figure 5). As mentioned above (Section 3.1), the system activities are performed in three steps. The citizen's question is processed in the first phase. This activity starts with steaming and elimination of stop words (question filtering). If the citizen selects the term(s), the question is added to the existing cluster specified by the selected term(s). Otherwise, ADVANSE performs a measurement of similarity between the new question and cluster centroids (cluster determining). By using Cosine similarity, ADVANSE compares the vector of question terms with the vector of key terms that belongs to the cluster (Section 4.5).

After the cluster is determined, ADVANSE starts the last activity in the first step. The engine gets the questions from the cluster, one by one, measuring the similarity with the new one. When the question is found, ADVANSE starts the second phase. If none of the clusters satisfy the threshold criteria, ADVANSE tries to find a document that is closest to the search criteria.

If a similar question is found, ADVANSE starts searching for the most appropriate answer. Questions, answers and their mutual connections are forming triplets. One question can be connected to many answers, but connections between them could have different connection strength. Connection strength and threshold are used for selection of an answer that best fits the question (answer finding). The answer with the highest connection strength will be selected if the connection strength is above the threshold. The last activity in the second step is content delivering (responding with a document or an answer).

In the last step, the system checks if there is citizen feedback. Feedback includes evaluation of the system response. ADVANSE processes feedback in two ways. If the feedback is related to the document, ADVANSE updates the weights of the document's key terms (Section 4.7), i.e. positive feedback increases the weight, while negative feedback decreases the weight. In the answer case, if feedback is positive, ADVANSE increases the strength of the Q & A relation and vice versa. Due to the term weighting mechanism, documents are clustered just one time (during initialization).

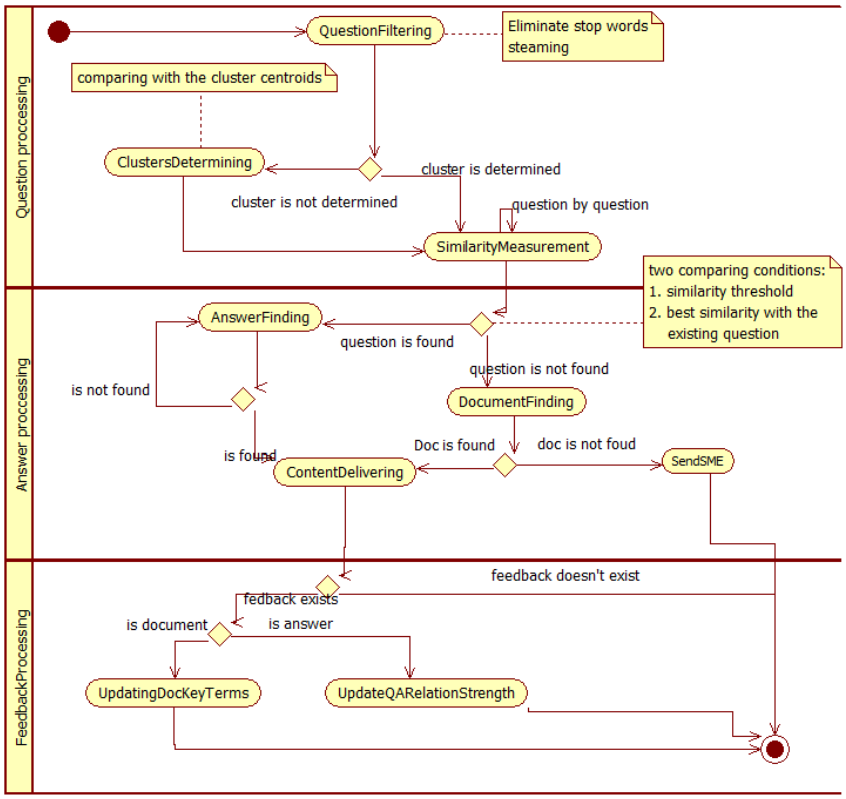


Figure 5
Processing algorithm

4.3 Concrete Knowledge Representation

In actual e – government systems, the SME suggests that citizens read answers to similar questions. This case is very common if there are a lot of questions and answers in the system. However, there may be one or more questions related to the same answer. If there is only one answer related to a question and the similarity between the cluster's questions and the new one is higher than the threshold value T, this answer is the only solution for a system response. However, this is not a common case. More often, the questions in the cluster can be related to different answers. Even more often, some questions could be related to more than one answer. The problem that arises here is how to select the appropriate answer to the given question.

The similarity can be functionally expressed as a maximum value of the relations between the questions in cluster C_p and the new one q_{new}(Equation 9).

$$sim(q_{new}, C_p) = \max_i \{sim(q_{new}, q_i) \mid q_i \in C_p\} \quad (9)$$

where $C_p = \{q_1, q_2, ..., q_m\}, i = 1,...,m,\ p = 1,...,s$

Based on this representation and the conceptual model, the following rule design is used in the system (Equation 10).

$$(\exists C_p)\,(sim(q_{new}, C_p) \geq Treshold) \Rightarrow \{a_j \mid \begin{matrix}(\exists t_{ij} \in T_p)(\exists q_i \in C_p) \wedge t_{ij} = (q_i, a_j, s_{ij}), \\ i = 1,...,m, \\ j = 1,...n\}\end{matrix} \quad (10)$$

where $T_p = \{(q_1, a_1, s_{11}), (q_2, a_1, s_{21}), ..., \{q_m, a_n, s_{mn}\}, i = 1,...,m,\ j = 1,...n\},\ p = 1,...,s$

The concept of triplets is introduced in the proposed solution. Every triplet is generally represented as a set $\{q_m, a_n, s_{mn}\}$ where q_m represents the question that belongs to the corpus C_p, a_n represent the answer related to the question q_m, and s_{mn} represent the strength of the relation between q_m and a_n. Searching for the most similar question can start if the cluster (C_p) that the new question belongs to is determined. There is an additional precondition for rule firing beyond similar questions in the cluster: the similarity has to be higher than the specified threshold. If both of these conditions are satisfied, the rule returns the answer related to question (q_m) that is the most similar to the new one (q_{new}). If multiple answers are connected with the selected question q_m, the one with the highest strength value will be chosen. However, other answers sorted by strength values will also be offered as alternatives to the recommended one in case the citizen is not satisfied with the offered answer.

The other consequence of searching is establishing new relation(s) between the new question and the answer(s) that the system replied with. The creating rules are designed for this purpose (Equation 11).

$$(\exists a_{res})(f(a_{res}) > 0) \Rightarrow (t_{new} = (q_{new}, a_{res}, s_{new}) \wedge T_{new} = T_p \cup \{t_{new}\}) \quad (11)$$

where a_{res} is responding answer, f is feedback function and t_{new} is a new triplet of a new question (q_{new}), the responding answer and new strength (s_{new}).

The new question is added into the cluster and the new question – answer – relation strength triplet is added to the knowledge base.

4.4 Measuring of Questions' Similarity

After the cluster is determined, ADVANSE uses the cluster's questions and compares them with the new one. Three similarity measures with different approaches are used for this purpose: *cosine* similarity, *Jaccard* correlation

coefficient and averaged *Kullback-Leibler* divergence. They are implemented as functions that are used by searching rules (Section 3.3). Common for all of the measures is that the question's text is represented as a set or a vector of terms and their additional properties such as term frequency, hash function, distribution of probability. A short description of the applied measures is presented in this section.

Cosine similarity is one of the most commonly used text similarity measures because of its simplicity and because it is not dependent on the text length. The compared texts are considered as resulted term vectors [9] and their similarity is expressed as a cosine of the angle between these vectors.

Jaccard Correlation Coefficient (JCC) is a similarity measure that depends on set theory [4]. Questions are considered a set of terms, and correlation between two of them is calculated as a ratio between the intersection and union of their sets. The zero value is calculated if there are not terms in the intersection. JCC has a maximum value (1) if both of the documents have the same term sets. The properties of terms are expressed by values of the appropriate hash function randomly selected from the hash function set (universal hash family for strings). If there is collision between hash values, the system performs rehashing. Different lengths of the questions are solved by padding the shorter document with zeros before comparison.

Averaged *Kullback-Leibler* (KL) divergence is a similarity measure that depends on probability theory [12]. Compared questions are represented by probability distributions of the terms they consist of. Because KL divergence is non – symmetric (the result depends on the order of comparison), averaging is used as a technique for compensation.

Using different similarity measures provides more scalability to the system. It is important because there are different domains (e.g. health, finance, low) in which these functions can be used. System performance can be evaluated during usage of different functions and the most appropriate function can be selected for searching purposes.

4.5 Q & A Relation Strength

Each time a new question-answer pair is created (see section 4.3) ADVANSE assigns an initial strength (V_{ini}) to their relation. This value is changeable and depends on a number of feedbacks and a feedback score (Equation 12). The relation strength is considered for ranking the answers related to a specified question.

$$S_{qa} = V_{ini} + I_f \cdot R_{pn} \tag{12}$$

The product of two factors: feedback importance (I_f) and positive & negative ratio (R_{pn}), represents the other part of the strength equation. It becomes important when the number of feedbacks exceed the threshold value. The feedback importance is the weighted factor included for this purpose. It is calculated on the normalized way (Equation 13):

$$I_f = 1 - \frac{1}{\sqrt{N_{pf} + N_{nf}}} \tag{13}$$

where N_{pf} is the number of positive feedbacks and N_{nf} is the number of negative feedbacks.

The importance value is greater than zero if the number of feedbacks is more than two. If the number of feedbacks grows, its value tends to be one. ADVANSE uses a threshold mechanism to define the number of feedbacks necessary for the strength calculation. In other words, if there are fewer feedbacks than defined by the threshold, the relation strength equals its initial value.

Another factor product of the relation strength is a ratio of positive & negative feedbacks (R_{pn}). It is calculated by dividing the sums of positive (M_p) and negative (M_n) grades with the whole number of feedbacks (Equation 14). The R_{pn} ratio is positive if the sum of positive marks is greater than sum of negative marks. If there is an equal number of positive and negative feedback, it has a zero value. Otherwise R_{pn} has a negative value.

$$R_{pn} = \frac{\sum_{i}^{N_{pf}} M_p - \sum_{i}^{N_{nf}} M_n}{N_{pf} + N_{nf}} \tag{14}$$

If the value of feedback importance (I_f) is below the threshold, the R_{pn} ratio is not included in the relation's strength calculation. Otherwise, the ratio value is included and decreased by the factor I_f. In practice, it is easier for citizens to get feedback by selecting one of two options than to evaluate the answer by selecting one of many grades. If there are only two feedback options (e.g. satisfied or not satisfied), the calculation of R_{pn} is simplified (Equation 15).

$$R_{pn} = \frac{N_{pf} - N_{nf}}{N_{pf} + N_{nf}} \tag{15}$$

In this case, the ratio's value is normalized in the range from -1 to 1. If there is not even one feedback, the relation strength is represented only with its initial value feedback (V_{ini}). Otherwise, the feedback importance (I_f) and ratio of positive and negative feedbacks (R_{pn}) are included in the calculation.

4.5.1 Responding with Documents

If the new citizen's question is not similar enough to any of the existing questions containing an answer, ADVANSE tries to find similar document(s) instead of answers. This measuring is performed in the same way as in the case of questions (Section 4.4). The documents are described by key terms that are statistically extracted during the clustering phase. They are clustered in the same way as the questions (Section 3.2) and they are presented as a set of key terms and their membership functions.

If the appropriate document is found (the similarity between the citizen's question and the document is greater than the threshold), ADVANSE returns a link to the document. Then, a relation between the question and the document is established. The initial strength is given and it is changeable over time depending on the citizens' feedback.

Conclusion

The represented hybrid solution depends on the nature of the e – government services. It is mainly focused on providing conditions for advanced responses to citizen requests. Most of the information that can be used are held in repositories as formal documents. Otherwise, citizens' questions and SME answers are recorded in the system DB. Therefore, the ADVANSE content model is layered. The content is fuzzy clustered based on fuzzy sets theory and fuzzy c-means algorithm. The boundaries between clusters do not exist and pieces of information can belong to more than one cluster. On the other hand, the questions, answers and documents are semantically connected in the system. The mentioned features have influence on the overall system design and they make the system flexible in responding to citizen queries. Using different text similarity measures provides adaptive behavior to the system. Processing of the citizens' questions in different and flexible ways provides the conditions for early high – quality responses. The expectation is that the citizens will be much more satisfied than before, able to make better decisions with better information, while the public administration will have captured more systematic information on the problems citizens face.

Improved e-government services in response time, quality of response (citizen's satisfaction) and the usage of existing content in a new manner as well as relaxing the SME responsibility for answering citizen questions represent the main results of the ADVANSE project. Adaptive response represents one of the most important features of ADVANSE. The questions on one side and the answers and documents on another are related. These relations are changeable and they depend on citizens' evaluation of the response (their satisfaction with the delivered content). Thus, the response to the same question can be different over the time as conditions and citizen needs change. Documents are delivered in response to a situation where there is not any suitable answer to the citizen question (if the threshold of similarity between the new and existing questions is not satisfied). ADVANSE

respondswith documentsthat containa key term set most similar to that of the question.

Future objectives will focus on testing in different environments and if necessary, to improve the ability of adaptation. The functioning of the system in multilingual environments will be the one of the targeted solutions. The dictionary and grammar of different languages enlarge the complexity of the system independent of the domain of usage. In this case, several processing strategies should be used. Functioning of the system in different e-government domains represents the other challenge. Different domains are covered with different thesauruses. More specialized similarity techniques will be required. Document processing will also need to be improved.Annotation (tagging) can provide a response with extracted part(s) of a document instead of the whole document body.

The presented solution is a part of a wider project aiming to provide an intelligent decision support system able to collect, cluster and analyze data from various data sources (social, biological, and economical systems) in order to make government decisions easier. Future research will take place in several directions, such as the improvement and evaluation of information retrieval and text mining algorithms, allowing personalized services by applying user profiles, implementation of morphology data of different languages for better text preprocessing and establishing semantically based relations between pieces of information.

Acknowledgement

This work was supported by the Ministry of Science and Technological Development of the Republic of Serbia under the grant III-44007.

References

[1] Garijo, F., Riquelme, J., Toro, M.: A GRASP Algorithm for Clustering, Proceedings IBERAMIA 2002, LNAI 2527, pp. 214-223, 2002

[2] Global E-Government Survey 2012, E-Government for the People, United Nations, New York, pp. 9-69, 2012 retrieved from http:// unpan3.un.org/egovkb/global_reports/12report.htm

[3] Hamerly, G. and Elkan, C.: Alternatives to the k-Means Algorithm that Find Betterclusterings. Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM), pp. 600-607, 2002

[4] Hamers, L., Hemeryck, Y., Herweyers, G., Janssen, M., Keters, H., Rousseau, R., Vanhoutte, A.: Similarity Measures in Scientometric Research: the Jaccard Index Versus Salton's Cosine Formula Information Processing and Management: an International Journal, Volume 25 Issue 3, pp. 315-318, 1989

[5] Bezdek, J., 1981, Pattern Recognition with Fuzzy Objective Function Algorithms, Kluwer Academic Publishers Norwell, MA, USA, 1981

[6] Laan, M., Pollard, K., Bryan, J.: A New Partitioning Around Medoids Algorithm, Journal of Statistical Computation and Simulation 73, No. 8, pp. 575-584, 2003

[7] MacQueen, J.: Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, pp. 281-297, 1967

[8] Robertson, S.: Understanding Inverse Document Frequency: On Theoretical Arguments for IDF, Journal of Documentation, Vol. 60, No. 5, pp. 503-520, 2004

[9] Salton, G. and Wang, A.: Generation and Search of Clustered Files. *ACM Transactions on Database Systems*, Vol. 3, No. 4, pp. 321-346, 1978

[10] Vattani, A.: K-means Requires Exponentially Many Iterations Even in the Plane, Discrete and Computational Geometry Journal, pp. 596-616, 2011

[11] Wu, H.C., Luk, R. W. P, Wong, K. F., and Kwok, K. L .: Interpreting TF-IDF Term Weights as Making Relevance Decisions, ACM Transactions on Information Systems, Vol. 26, No. 3, Article 13, 2008

[12] Zhang, W., Ma, J., Zhong, Y.: UsingKullback-Leibler Divergence Language Models to Find Experts in Enterprise Corpora, IITAW '09: Proceedings of the 2009 Third International Symposium on Intelligent Information Technology Application Workshops, pp. 402-405, 2009

[13] Bezdek, J., Enrlich, R., Full, W., FCM: The Fuzzy c-means Clustering Algorithm, Computers & Geoscience, Vol. 10, No. 2-3, pp. 191-203, 1984

Styles or Cultural Background does Influence the Colors of Virtual Reality Games?

Cecília Sik-Lányi

University of Pannonia, Egyetem u. 10, H-8200 Veszprém, Hungary
lanyi@almos.uni-pannon.hu

Abstract: Illustrated color of known objects in Virtual Reality (VR) games for various games exhibit characteristic differences. In the first part of our analysis we studied the cartoons' colors, with a great emphasis on whether there are any cultural differences. Coloration of well-known objects depicted in cartoons originating from different parts of the world show characteristic differences. We analyzed several soft-copy and hard-copy cartoons form all over the world and determined what colors the designer use for complexion, sky, water, soil, etc. We continued with the study of VR games' colors. We analyzed eight game categories' colors and determined which colors the designer used for skin, sky, water, grass, etc. These colors were compared with the prototypical memory colors and cartoon colors of these objects. The research quantifies these differences and provides advice to the VR games to be tinted, if they are intended for a specific region of the world and specific VR games. In the second phase of the research such images of films, which were the equivalent of VR games are analyzed. The staining of these films were compared to the corresponding color of VR games and memory color display object.

Keywords: Virtual Reality; Game; Film; Color

1 Introduction

Many books and articles deal with the question of how colors influence the mood of people seeing them and how, e.g. in a picture the mood of a person can be expressed in colors. For artists, color was always a vehicle to express moods [1]. Panton, as an artist, even gave the title of his booklet: "Choosing colors should not be a gamble. It should be a conscious decision. Colors have meaning and function" [2]. Hutchings [3] discussed the use of colors during the ages, and pointed out that there are cultural differences that should be taken into consideration. Robertson and his colleagues [4] found evidence of cultural and linguistic relativity, among others in color categorization. Analyzing ninety-eight languages Berlin and Kay [5] found that eleven color words act as focal points of all the basic color words in all languages of the world. This set of eleven seems therefore to be a semantic universal. Basic color words are translatable. These

basic color terms are in English: red, orange, yellow, green, blue, purple, pink, brown, grey, black and white.

Color perception has been a traditional test-case of Whorf's principle of linguistic relativity [6], [7], [8], i.e., the idea that speakers of different languages perceive and process reality and the world differently, influenced by lexical and grammatical distinctions specific to their language. The vast majority of empirical research in the past 17 years has supported the notion [9] that language acts an attention-directing mechanism in the cognitive processing of color, in both offline similarity judgments [7], [4] and online perceptual discrimination [10], [11], [12].

Duncan and Nobs [13] investigated the interrelationship between human emotions induced by colors and their psychophysical stimuli, and found differences between emotional color scales established in Europe and the Far East. Szabo developed predictive mathematical models of color harmony [14] to quantify color harmony impression of observers and a new light source quality metric called Harmony Rendering Index [15] based on these formulae.

Multimedia applications usually use graphical drawings instead of photographs, because they can be more efficiently moved and stored. In many of the cases users regard pictures with less colors appropriate for use. For VR games' simulations it is enough to use homogenly designated pictures, like in the programs for treatment of some kinds of phobias. These pictures will henceforward be called cartoon-like pictures, graphical pictures or simply cartoons.

The graphical designer has to choose from a great amount of hues representing qualities. Most of the multimedia programs enable a wide range of choice of colors from its own palette, and also gives some instructions how and where these colors can be used. Previously the coloring of the pictures was the task of the graphical designer, who learnt his profession. Today, this task is made by the IT engineer programming the animation, or more frequently called animator.

The pictures have an aesthetic value as well, which means that the chosen hues cannot clash and that they have to be in a harmony. There are several guiding principles regarding the harmony of colorization.

In our research we give instructions how to set the colors of different qualities (objects). We created a database, which contains hundreds of pictures' colors from different cultural regions. We categorized hundreds of pictures that are paper-based (henceforth hard-copy) or pictures from the internet (henceforth soft-copy) and also we measured the colors of their different qualities. We evaluated six important and most frequently used qualities (face, grass, sky, lake, foliage, tree-trunk) from four different cultural regions. We compared the results with the memory colors.

The above investigations were performed either on single color patches or tried to elaborate on historic findings. We were seeking a different way to be able to compare the preferred coloration of well-known objects by present day population

– especially young people – coming from different cultural backgrounds. VR games are popular among children and young people all over the world. They are not only popular, but have proved effective when used in special education to teach independent living skills [16], [17], and more latterly by 'modding' popular games engines (such as the Source engine as used in Half-Life 2) to teach employment skills [18], [19].

A lot of 3D games can be found nowadays. There is a huge difference between the properties of game heroes and the properties of real humans. This difference can be seen in the choice of colors to depict the heroes and their surroundings, which is far from what we can see in our everyday life and our environment. The usage of these computer games by children is getting longer and longer every day. The main question can be, whether the colorization of these games have any influence on the children's color sense? If we want to answer this question we have to analyze the input, the colors which are used in games.

Studying cultural differences that impact humans' interaction with information is an emerging field [20], [21, [22]. Inside this field, our research focus on cultural differences of colors. It is known that color may influence human emotions or feelings, in the sense that some colors may make one happy, while some colors may make one depressive [23]. Sato and co-workers split color emotion into three categories (activity, potency and warm-cool) in Japan [24]. Analysis of cross-cultural color emotion was investigated in 7 countries [23]. In this study they evaluated 214 color samples on 12 emotion variables by subject from seven different regions in a psychophysical experiment. By factor analysis, it was found that three factors were sufficient to represent 80 "region-emotion" variables. They found that chroma and lightness were the most important factors on color emotion.

Virtual Environment (VE): A synthetic, spatial (usually 3D) world seen from a first-person's point of view. The view in a VE is under the real-time control of the user. Virtual Reality (VR) and Virtual World are more or less synonymous with VE [25]. More specifically, VEs are distinguished from other simulator systems by their capacity to portray three-dimensional (3D) spatial information in a variety of modalities, their ability to exploit users' natural input behaviors for human-computer interaction, and their potential to "immerse" the user in the virtual world. The effects of human differences in immersive VR environments are a cutting edge research topic [26, 27]. Inside this area, focusing on cross-cultural aspects is our promising target.

According to Steinkuehler, the current global player populations of the most popular three games that she has studied over the past few years, totals over 9.5 million - a population which rivals, e.g. most US metropolises [28].

These games are used on PC or laptop. In this case there are non immerse VR games. The users observe these games under approximately 30° visual angle, thus in this respect it is immaterial that people do not see colors outside of approximately 100° horizontally [29]. Another question is if the player uses these

games using a Head Mounted Display (HMD). Field of view (FOV) of the HMD gives the world builder yet more compromises to make. Although theory is limited, narrow FOVs may hinder task performances such as maneuvering, grasping objects and locating moving targets [30]. Wider FOVs may improve performance and also feelings of involvement and presence, but this comes at the expense of greater weight and size of the HMD and possibly worse image resolution [31].

Virtual reality games are popular among children and adolescents around the world. The coloring of the heroes and the environments in VR games are very far from the average people and the real environment. Children play VR games every day, so these games affect the aesthetics of the children. The question would be: Do not pay more attention to children experiencing right discoloration of virtual worlds: in this context? In order to provide designers and programmers to pay due attention to these colors in the virtual world, or not?

2 Colorimetric Fundamentals

2.1 Color

In scientific literature three components of color perception are distinguished: brightness, hue and colorfulness:

- Brightness: the sensation can be almost blending strong, medium or dim and dark.
- Hue is usually shown as a hue circle, where four distinguishable different areas are red, yellow, green and blue.
- Colorfulness has been divided by MacDonald into a five value scale: Starting with gray (achromatic) up to the most brilliant color.

The communication of the color perception can thus be made in the following form: The complexion of the person who stands in front of me is medium bright, moderately vivid and reddish yellow [32], [33].

The color stimulus can be described in a definite way using a color order system. There are several color systems in use RGB, CMYK, CIELAB. We performed our measurements using the CIELAB system, it is the system now recommended by international standards [34]. Figure 1 shows this color system as a three dimensional body, with the lightness, chroma axes and hue circle (or a^*, b^* axes to describe chroma and hue) where the hue is measured as a hue angle. It provides reasonable uniform color scales in lightness (L^*) and chroma (C_{ab}^*) and gives more or less equidistant scaling along the hue circle (h_{ab}^*).

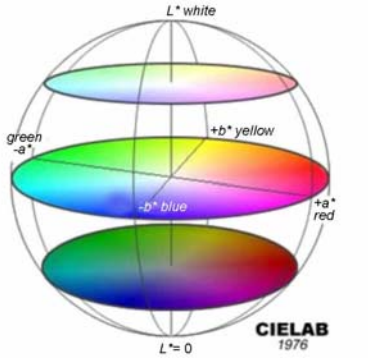


Figure 1
Three-dimensional graph of the real surface colors, in the system of lightness, chroma and hue

2.2 Memory Colors of Well Known Objects

The term memory color is used for the color of well-known, often seen objects, as in our brain we attach a color to the given object. Memory colors are well stabilized products of our memory. They are colors we will pick from a high number of color chips if one is asked to show the color chip resembling the color of human complexion, or sky blue, etc. Table 1 shows the *L*, a*, b** values of some memory colors [35].

Table 1
Memory colors of well-known objects according to the different authors

Memory color	L*	h_{ab}*	Reference
Caucasian skin	79.5	32.9	[36]
blue sky	54.0	238.8	[36]
green grass	50.0	138.5	[37]
oriental skin	63.9	49.0	[38]
deciduous foliage	33.6	145.3	[36]

3 Measuring Method

3.1 Cartoon's Picture Samples Studies

In our investigations we had to restrict ourselves to pictures where the theme of the picture has some resemblance to the real world, i.e. pictures were not considered, where the coloration was far from what one would accept as a

"natural" color of an object. For representative pictures and more detailed measurement results see the WEB site of one of the authors

With the start of our research we posed the question that where can we find hundreds of pictures that can appear everywhere in the world? How can we map the pictures that were created by graphical designers during the last decades? The solutions are the cartoons, both in hard-copy and soft-copy. As a first step we collected hundreds of graphical pictures (henceforth we call them cartoons). To reach our aim we tried to categorize them according to cultural regions. The most important categories are the following: European (historical), American (hero, combat), Japanese (so called manga) and Australian (family).

The European carton designers love to depict historical themed stories, here we have also listed the depiction of literary works and situation comedies. These are stories like the French "Asterix and Obelix", or the Hungarian "Mátyás király", "A Pál utcai fiúk", or the English "Tom Sawyer or Huckleberry Finn", etc. Regrettably the real Hungarian cartoon designers have disappeared, however we were able to find a web site (http://rajz.film.hu/), where they wish to restart making the traditional cartoons with a help of a project.

The Americans like completely different types of stories. These are called 'combat' cartoons (Spiderman, Robocop, Superman). In an American cartoon the depiction is completely different from the one we know in Europe. The figures are sketchy; we can feel the movement and dynamics on them. It is interesting that in these cartoons are the facial and body forms depicted in the most natural way. The scenery however is not as important, mostly they depict a figure with a simple one color background. Despite these the Europeans give the background nearly the same prominence like on the characters.

The cartoons which are preferred by the Japanese are the most interesting. The Japanese cartoon designers do not present their characters on the basis of their cultural background. The characters, who are mostly children, are tall, have a long leg, big and colorful eyes and hair, and have a really pale complexion. On the basis of the previous knowledge we can unequivocally divide these cartoons from the previous categories.

We got some sample from the Australians as well, but we could not really categorize them. So we created a new category, based on the Australian type (family). The Australian cartoons' characteristics are nearly the same as we can find in the European or the Japanese types, the figures and the background are clearly distinct, but mostly the scenes are mostly in the present, they are fabular, the colors are natural, there are no unnatural hues, green or purple hairs.

There are cartoon series that have a global distribution, and where prints are made in different countries. One of these is 'Asterix' that has more the 100 translations. Just to get a feel on the differences publishers choose for complexion color we summarized complexion color of "Cleopatra", one of the cartoon characters in the

series No. 6 “Asterix and Cleopatra” depicted in the Cover Galery of Asterix International in Figure 3. We just inserted the pictures in a graphics program that enabled “eye-drop” technique to fetch the color into the program’s color management. As we were interested only in the relative differences between the particular editions, we used the default setting of the program. Table 2

shows the results in increasing CIELAB lightness. As can be seen the Swedish and the American editions use low CIE 1976 lightness, in case of the American edition extremely high CIELAB chroma was found. The next higher lightness was found for the Turkish edition, where the CIELAB hue angle is at the reddish extreme. The Hungarian, Russian, Greek and German editions show complexion colors of higher lightness and low CIELAB chroma. The latter two are also between the most yellowish ones.

As we have downloaded the above samples from the Internet, we have no information on eventual distortions produced by the scanner at the input site, neither on the actual state of the pages scanned. Therefore we do not want to draw major conclusions from this part of the study; it should show only that even for the same fundamental picture if reproduced in different countries color differences would be seen that might be coupled to the regional preference. To get a better insight into the regional differences we investigated cartoons received from different parts of the globe, and compared the coloration used for a number of representative objects.

Table 2
CIELAB co-ordinates of some complexion colors of Cleopatra in the Asterix and Cleopatra series of the Asterix International Cover Galery

Edition	L^*	C_{ab}^*	h_{ab}^*
Sweden	64	76	42
USA	67	107	46
Turkey	67	71	37
Brazil	70	74	46
Italy	70	76	42
Korea	71	77	48
United Kingdom	74	62	52
The Netherlands	74	84	48
South-Africa	76	64	47
Hungary	78	51	42
Russia	82	50	47
Greece	83	52	67
Germany	85	48	56

3.2 The Investigated Representative Hard-Copy Cartoon Colors

Cartoons have been received from Australia, Europe, Japan and the USA. With the help of a small CCD array based spectroradiometer we determined the spectral reflectance properties of small homogeneous patches (approximately 3 mm diameter) using incandescent lamp illumination and 45°/0° measuring geometry. From this CIELAB values were calculated for D65 illuminant and 2 degree standard observer.

Analyzing a high number of cartoons we determined the coloration of the following objects:

- fair, suntanned and dark complexion,
- black, blond, brown and red hair,
- clouds and sky,
- tree trunk, grass, foliage,
- soil, sand and water (lake),
- concrete.

In the following data of six object colors will be detailed: complexion, sky, tree trunk, grass, foliage and water (lake). For better visualization Figure 2 shows average values and standard deviation ellipses for three-three of the analyzed six colors.

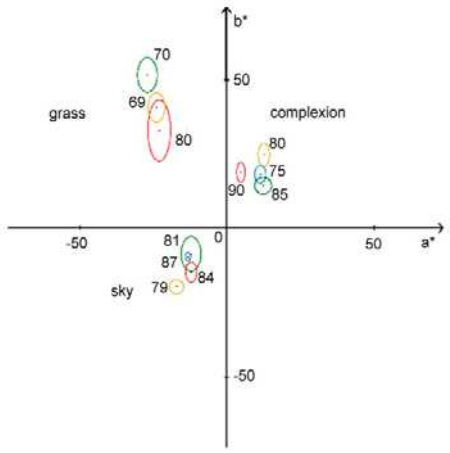


Figure 2

Typical complexion, grass and sky CIELAB values, depicted on an a^*, b^* diagram, L^* values are written in the vicinity of the standard deviation ellipses, shown in blue for American, in green for Australian, in red for Japanese and in yellow for European hard-copy cartoons.

As an example we can see one of the edition of *Asterix*'s (original) cover (Fig. 3).

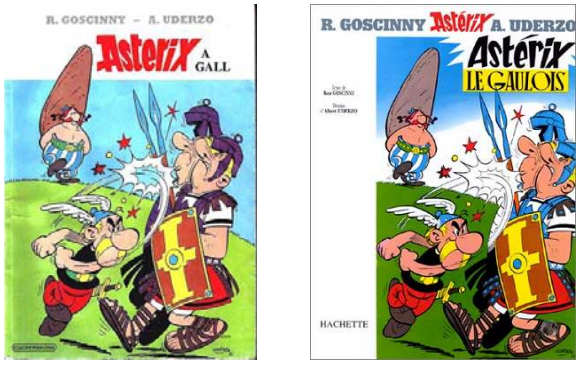


Figure 3
Cover released in Hungary (left side) and cover from France (right side)

3.3 Representative Soft-Copy Cartoon Colors

Similar investigations as described in the previous paragraph were performed also on cartoon pictures found on the Internet. The only difference in this case was that the CIELAB co-ordinates were not measured in our laboratory. We supposed that the cartoons were put onto the Internet using sRGB color space, and thus we have set our graphics program to this default state and determined the CIELAB co-ordinates using the eye-drop facility of the program. Here again we show in Figure 4 average CIELAB values and standard deviation ellipses for three-three of the analyzed six colors.

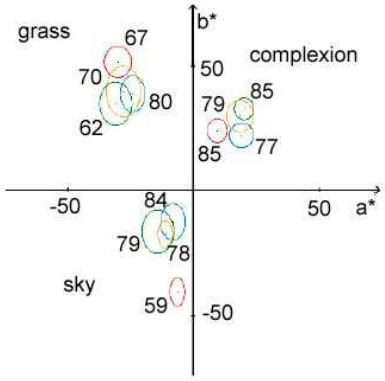


Figure 4
Typical complexion, grass and sky CIELAB values, depicted on an a*, b* diagram, L* values are written in the vicinity of the standard deviation ellipses, shown in blue for American, in green for Australian, in red for Japanese and in yellow for European soft-copy cartoons.

Table 3
Numerical data of hues of objects on the Hungarian and French covers

Object		*L**	*a**	*b**
Grass	HU	80	-36	59
	FR	50	-36	52
Sky	HU	88	-27	-18
	FR	69	-10	-46
Face	HU	87	12	23
	FR	73	34	35
Rock	HU	57	34	22
	FR	57	15	4
Moustache	HU	96	-14	79
	FR	88	-3	82

In Table 3 we can see the differences between the two pictures' hue usage. One of the most noticeable differences is that the Hungarian release uses lighter colors. During the analysis of the grass color there is not a great differences between the *a** and *b** value. However the other values show some characteristic differences, the "Hungarian" sky is greenish, the rock is brownish. The French face is reddish and darker.

According to the data shown above, we can assess, that we can get realistic results from the measurement of the hues of objects, if the models were made in equal conditions. This was the reason why we collected cartoons from other parts of the world. We succeeded to collect some cartoons from Australia, Japan, Korea, America and France. Unfortunately there are no cartoons on the market that has been drawn and printed in Hungary. We did not dare to analyze twenty year old cartoons, because the quality of paper and the printing technologies have changed a lot through the last decades, and our results could not be compared with the results from the cartoons, that we got from the other countries. To conclude, we only measured the parts of pictures from original cartoons. This way we would also like to thank the people, who were so kind as to send us some cartoons from abroad.

3.4 Measuring the Colors of VR Games

Most popular virtual games were categorized into eight groups, as can be seen in Table 2. Pictures have been downloaded from the Internet of 89 VR games: 7-10 pictures from every game, altogether 752 pictures. Movie films, corresponding to the above films have also been analyzed, 179 pictures from 20 films have been downloaded from the internet. The category of games we used in our research is shown in Table 4.

Table 4
Categorization of the most popular virtual games

Name of the game category	Number of the game category	Number of the film category
Action, Adventure, Mystery Games	G1	F1
Children's Games	G2	F2
Driving & Racing	G3	-
First-person Shooters	G4	F4
Simulations	G5	-
Role-playing Games	G6	F6
Strategy	G7	-
Sports	G8	-

3.5 Games Categories

Title of games, from which pictures were taken can be seen as enumerated below.

Action, Adventure, Mystery Games:

Tomb Raider 7
Silent Hill 4: The Room
Alone in The Dark: The New Nightmare
GTA San Andreas
Resident Evil 4
Legend of Zelda: The Twighlight Princess
Myst V: End of Ages
Syberia II
The House of the Dead III
Monkey Island 4

Children's Games:

Fame Academy: Dance Edition
Camgoo + Webcam
Shrek 2 Team Action
Jimmy Neutron Boy Genius
Robots
Spongebob Squarepants: Battle for Bikini Bottom
Harry Potter
Quidditch World Cup – Classic

Driving & Racing:

Need For Speed Underground II
Nascar 2005: Chase for the Cup
Colin McRae Rally 2005
TOCA Race Driver 2
Gran Turismo 4
Driv3r
Hot Wheels Stunt Track Challenge

First-person Shooters:

Counter-Strike Source
Day of Defeat Source
Battlefield 1942
Medal of Honor Pacific Assault
James Bond 007: Nightfire
Call of Duty
Doom 3
Unreal Tournament
Maxpayne II
Star Warp Rep. Commando

Simulations:

Sim City 4
The Sims 2
Microsoft Flight Simulator 2004
Animal Crossing (Game Cube-ra)
Will of Steel
Empire
F/A-18: Operation Desert Storm
Pacific Fighters
IL-2 Sturmovik - Forgotten Battles

Role-playing Games:

Final Fantasy X – XII
World of Warcraft
Lineage II
Ragnarok Online
Ultima Online Samurai

Guildwars
Chrono Cross

Strategy:

Cossacks II: Napoleonic Wars
Warcraft III (The Reign of Chaos
- The Frozen Throne)
Heroes of Might and Magic III – IV
Rome: Total War
Blitzkrieg II
Age of Mythology
Warlords 4
2005
Warhammer 40k
Imperial Glory

Sports:

Ski Racing 2005
MVP Baseball 2005
NHL 2005
FIFA 2005
NBA Live 2005
Fight Night Round 2
Madden NFL 2005
Tiger Woods PGA Tour

3.6 Film Categories

Action, Adventure, Mystery:

Lara Croft: Tomb Raider
Lara Croft and the Cradle of Life:
……Tomb Raider 2
Silent Hill
Alone in The Dark
Resident Evil
Resident Evil: Apocalypse

Children's:

Shrek
Shrek2
Robots
Harry Potter and the Sorcerer's Stone
Harry Potter and the Chamber of
Secrets
Harry Potter and the Goblet of Fire
Harry Potter and the Prisoner of
Azkaban

First-person Shooters:

James Bond:The World is Not Enough
James Bond: Die Another Day
Doom
Star Wars: Episode I - The Phantom Menace
Star Wars: Episode II - Attack of the Clones
Star Wars: Episode III - Revenge of the Sith

Role-playing:

Final Fantasy: The Spirits Within

3.7 Samples Studies

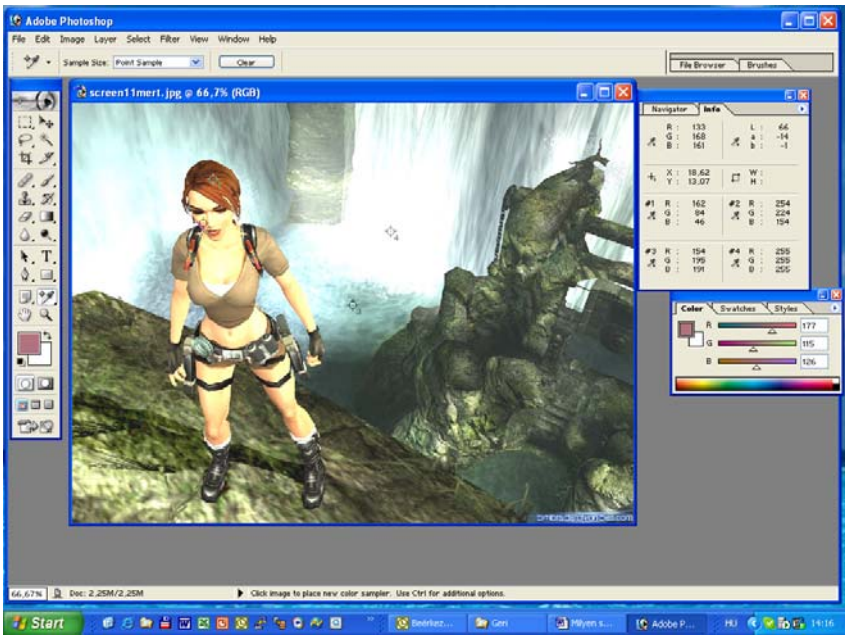


Figure 5

Color measurement in Photoshop, using the eye-drop tool, example is the Tomb Raider play.
1. measurement point: red hair, 2. measurement point: bright face,
3. measurement point: water, 4.measurement point: whitest point in the picture.

We selected objects to be measured for example skin (caucasian face skin, african skin), hair, sky, grass, trunk, cloud, water (lake, river, sea), brick, concrete (road) and so on.

During the investigations we wanted to stay within the bound of theme of pictures, where we have some resemblance to the real world, i.e. pictures where the coloration would not be accepted as a "natural" color of an object, were not considered, i.e. where a specific state of the mind of the hero (e.g. greed, anger, etc.) were emphasized by the designer of the game.

The pictures from the virtual games, which we analyzed, were downloaded from the Internet, supposing that these pictures were put on the Internet using sRGB color space. Our measurements were taken in the CIELAB color space. Sample Tool of Adobe Photoshop was used for sampling (Figure 5). Of L^*, a^*, b^* values nearly 4500 determinations were made. h_{ab} hue angle and its standard deviation (Δh_{ab}) was calculated (h_{ab} = arctan (b^*/a^*)), together with the chroma (C_{ab}) and its standard deviation, from the L^*, a^*, b^*, which enabled to determinate the changes in dimensions near to those of perceptions.

Figure 6
Picture's colors measured of Lara Croft Tomb Raider film

Tables 6 and 7 show the average CIELAB L^*, a^*, b^*, h_{ab} and Δh_{ab} values by game and film category. (Where the value is missing, there was no appreciable sample.)

Table 6
The average CIELAB L^*, a^*, b^*, h_{ab} and Δh_{ab} values by game category

Measured objects	Game Category	L^*	a^*	b^*	h_{ab}	Δh_{ab}
Caucasian face skin	G1	65.12	13.32	21.83	60.82	15.80
	G2	76.67	14.00	25.38	73.38	64.90
	G3	55.00	26.00	25.00	43.88	
	G4	51.44	13.56	20.29	57.23	21.86
	G5	71.00	22.00	31.00	54.80	4.72
	G6	69.93	18.93	26.67	56.49	9.07
	G7	83.00	12.00	35.00	71.08	
	G8	57.18	18.59	19.91	46.43	10.09
	G1	59.53	-1.71	-12.43	221.30	74.10
	G2	60.17	-7.91	-33.39	253.03	27.88
	G3	54.60	-3.08	-22.66	257.56	23.92

sky	G4	65.17	-2.37	-11.03	226.39	75.98
	G5	67.42	0,36	-29.39	270.51	8.05
	G6	61.12	-0.68	-31.04	266.48	17.88
	G7	70.95	-5.32	-22.14	256.80	17.76
	G8	56.00	-1.10	-28.30	265.52	18.25
grass	G1	44.47	-9.60	28.57	108.40	15.97
	G2	38.26	-25.42	32.26	128.17	8.91
	G3	42.92	-6.04	23.04	105.62	9.79
	G4	38.79	-4.07	21.93	101.27	8.40
	G5	43.57	-15,50	26.86	116.29	15.69
	G6	36.25	-12.85	28.80	114.55	11.82
	G7	41.73	-13.84	32.90	112.45	12.19
	G8	44.85	-17.37	32.46	118.74	8.89

Table 7

The average L^*, a^*, b^* coordinates as well as the calculated h_{ab} and its scatter Δh_{ab} values for the most important six parts of the picture for each film which has a corresponding game

Measured objects	Game Category	L^*	a^*	b^*	h_{ab}	Δh_{ab}
Caucasian face skin	F1	67.09	17.09	19.27	64.85	70.77
	F2	69.08	15.00	15.62	65.97	71.85
	F4	58.05	17.00	23.30	52.91	12.17
	F6	50.57	13.71	12.71	78.96	92.58
sky	F1	95.00	3.00	-5.50	281.12	53.30
	F2	70.23	-5.08	-28.38	241.57	54.74
	F4	70.78	-2.78	-13.89	224.82	59.94
	F6	71.67	6.33	-11.67	295.82	22.71
grass	F1	37.67	-2.33	15.33	99.01	5.08
	F2	52.47	-13.47	39.20	110.22	15.84
	F4	74.00	-19.00	49.00	111.19	
	F6					

4 Discussion

It seems that comparing the choice of colors used in virtual games [39] with the colors used in cartoons and the so called prototypical colors, which people mentally link with the colors of some objects, are of great interest. We compared results of studies by Sik Lanyi and coworkers [40], who investigated the usage of color shades in cartoons all over the world, with the colors of objects in our research, and with the ones determined by Tarczali [37] and Bartleson [36] for

memory colors. Sik Lanyi [40] found that there are characteristic differences between both of hard- and soft-copies (i.e. printed cartoons and downloaded images from the Internet) of cartoons originating from different parts of the world.

Complexion color: The lightest of all complexions is the Japanese, followed by the Australian. The darkest complexion color is the American both in soft- and hard-copy. Interesting facts is that the Japanese uses the palest and the least reddish color scale of all. We can observe that the memory color is much more pinkish than the cartoons use for depicting complexion colors. For depicting a sun tanned complexion we could find nearly the same tendency, where the Japanese is the lightest and its chromaticity is closely related to the European one, and the Australian is the pinkest. The memory color is the darkest. Average games use a more yellowish color than the cartoons and the memory colors. The most beautiful face colors were found to be the G2, G5 and G7 face colors. The others use colors that are far from the memory colors. The colors which correspond to the natural complexion colors are used in G2, G5 and G7 games, while the other games use a scale, which are very far from memory colors.

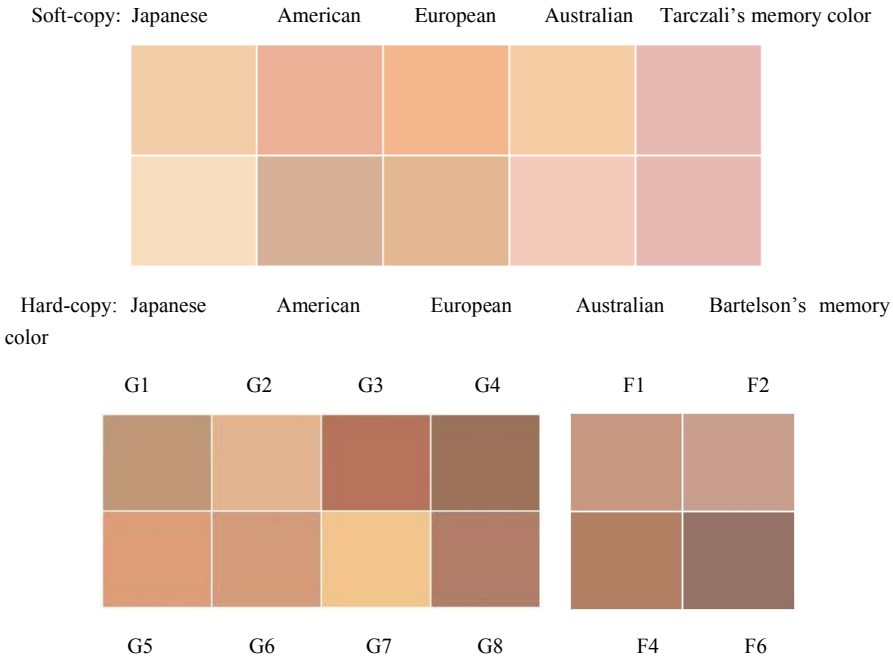


Figure 7
Comparison of the Caucasian complexion colors

In the films much larger scatter was found among the different samples, but the average values did not deviate much from the game samples, in most cases they were slightly lighter, an exception was the F6, where the average value of the Caucasian complexion color turned out to be much darker than in the games. As we had only one such film in our database, with some very special scenes, from this no direct conclusion can be drawn (Figure 7 and Figure 10-11).

Grass color: As for the grass color we can see several differences as well. The soft-copy grass color of the Japanese is much darker than the one of the hard-copy. The American or the European do not use such a yellow shade as the Japanese and the Australian. The memory colors, as well as the cartoon colors are all lighter determined in the mentioned two groups. The average colors of the grass in the virtual games are shown in the lower row. The only acceptable grass colors are G2 and G8. The others are more brownish than the memory colors. As for the film colors, most of the grass and foliage samples were of background values, which produced in the F1 category very dark colors, but e.g. in F2 and F4 films the colors of the grass correspond to the real life grass shade (notwithstanding the colors in games, the lightness was higher, resembling the real grass hue, which is, despite the memory color, more yellowish- Figure 8 and Figure 10-11).

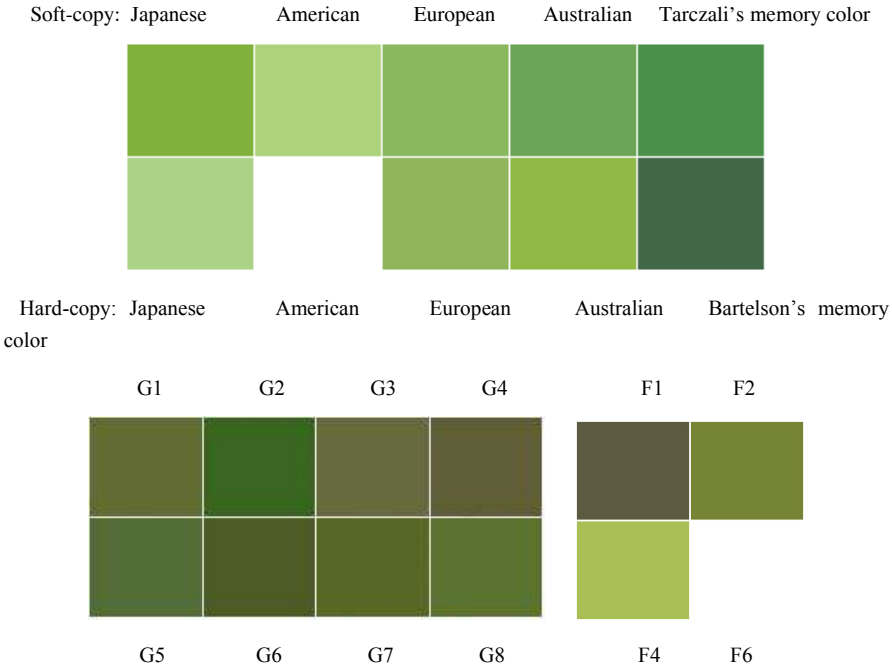


Figure 8
Comparison of the grass colors

Sky color: In the case of sky colors the cartoon designers use a much more lighter hue, than the memory color. The only exception is the Japanese soft-copy, which uses a quite dark shade. The sky colors used at the G1 and G4 are very grey. Films usually use a brighter color than the equivalent games, which is in the case of F1 almost white. In F6, the hue angle turned out to be extremely violet (Figure 9 and Figure 110-11).

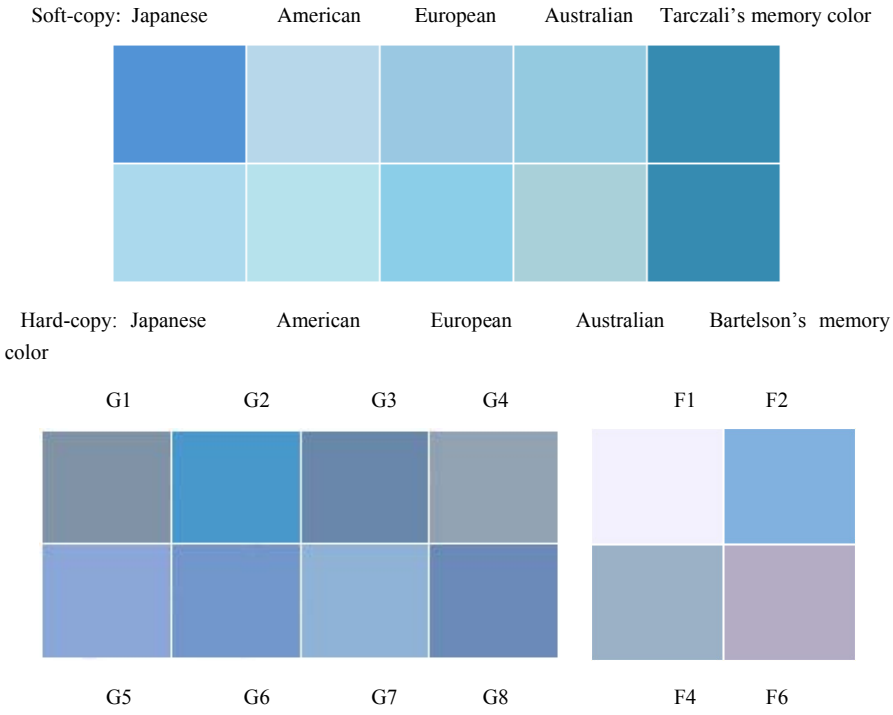


Figure 9
Comparison of the sky colors

Summary and Conclusion

Characteristic differences were found between the uses of colors for well-known objects in the different global regions. Designers can use these results if they have to prepare e.g. multimedia presentations for different observers.

In Japanese cartoons prepared for Internet presentation, stronger, more saturated colors are used, sometimes with lower lightness. In Europe the paler colors are preferred. This is true for most objects, except for complexion, where the Japanese use the palest colors and but it is interesting that the hue angle is larger than the one found in European complexion colors.

The situation is different in the case of printed cartoons. In print the Japanese use paler colors, and the most vivid colors are found in Australian pictures. American cartoon artists do not often try to use colors that resemble those of real life objects.

The prototypical memory colors are in most cases much darker then the colors used in the cartoons.

Different virtual reality games show some characteristic differences in the use of colors for some well-known objects. Our results can be presented with the help of some pictures. We took a picture from a book for outline drawings for painting, where the objects, whose colors were investigated (wood, grass, sky, etc.) were available, and painted it with Photoshop, to help the visualization of the colorimetric data. We colored these pictures according to the different game categories (Figure 10) and film categories (Figure 11) as well.

Looking at the pictures one has to consider the following: we used the most usual sky color for painting the sky, cloud color for clouds, tree trunk color for tree trunks, foliage color for leaves of bushes and trees, grass color for grass, and also the little pond on the lower right corner was painted with the average water color for every given game type. The girl on the left side of the picture was painted with the Caucasian complexion color, and the girl on the right side with the complexion color found for Afro-American persons was used. (Objects left white in the pictures do not serve enough examples to investigate.) In the cases of G1 and G4 game categories it is very interesting, that the color of the water in the right lower corner brings to mind the color of concrete road not of water, the sky is rather grey than blue, and the colors of grass and foliage are basically of same hue. Observe also the differences in foliage and grass as well as soil, sand, sky and water colors.

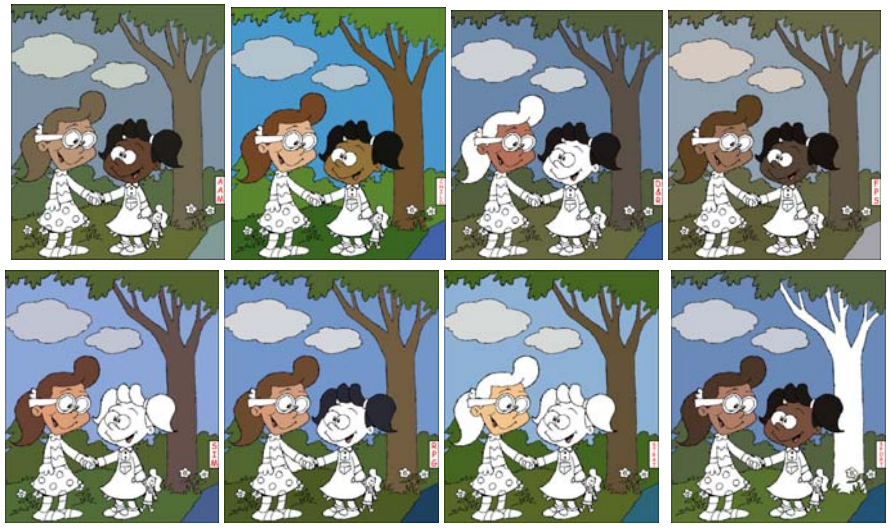


Figure 10
The same picture colored according the style used in different game categories: upper range: G1,G2,G3,G4, lower range: G5,G6,G7,G8

Inadequately, the designers of virtual games do not take care of using natural colors, which are far from the memory colors too.

During the investigations we found out, that designers use colors to emphasize the message of scenes, and the heroes' state of mind, and they forget about finding and using colors which harmonize with the colors of real world.

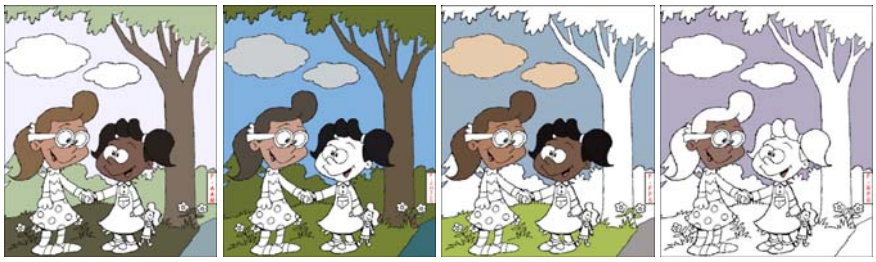


Figure 11
The same picture colored according the style used in different film categories: F1,F2,F4,F6

To make it perfectly clear, we would like to add, that in these virtual games we see fake colors [39], [40]. We would also like advise the designers of virtual games to use more natural shades of colors, which are more related to real ones.

Acknowledgements

The author thanks the help of her sons Andras Sik and Gergely Sik to download the pictures of the games and helping in the measurements. We want to thank for Veronika Végh, who has collected the cartoons and has helped us with the measurements.

References

[1] Itten V., Kunst und Farbe, Otto Maier Verl. Ravensburg, 1970

[2] Panton V., "Choosing colours should not be a gamble. It should be a conscious decision. Colours have meaning and function" as a title of his booklet, Danish Design Centre 1997

[3] Hutchings J.B., Color in anthropology and folklore, in Color for science, art and technology, ed. by: K Nassau, Elsevier, Amsterdam, 1998

[4] Robertson D., Davies I., Davidoff J, Color categories are not universal: replications and new evidence from stone-age culture. *J. of Exp. Psychology*: 2000, General **129.** 369-98

[5] Berlin B., Kay P., Basic Color Terms. Their Universality and Evolution, Berkeley: University of California Press, Reprinted 1991

[6] Heider E.R., Oliver D.C., The structure of the colour space in naming and memory for two languages, *Cogn. Psychol.*, 1972; **3**:337-354

[7] Davidoff J., Davies I., Roberson D., Colour categories in a store-age tribe. *Nature*, 1999; **398**:203-204

[8] Brown R.W., Lenneberg E.H., A study in language and congnition, *J Abnorm Psychol.* 1954;**49**:454-462

[9] Thierry G., Athanasopoulos P., Wiggett A., Dering B., Kuipers J-R., Unconscious effects of language-specific terminology on preattentive color perception, *Proc Natl Acad Sci USA,* 2009;**106**(11): 4567-4570

[10] Winawer J. et al. Russian blues reveal effects of language on color discrimination. *Proc Natl Acad Sci USA*, 2007; **104**:7780-7785

[11] Gilbert A.L., Regier T., Kay P., Ivry R.B., Whorf hypothesis is supported in the right visual field but not the left. *Proc Natl Acad Sci USA*. 2006; **103**:489-494

[12] Drivonikou G.V., et al. Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proc Natl Acad Sci USA*. 2007;**104**:1097-1102

[13] Duncan, J.,Nobs, H.J.,"Coloring our emotions: the measurement and application of our responses to color", Abstract. PICS 04 Conference Glasgow, 2004

[14] Szabó F., Bodrogi P., Schanda J. Experimental Modelling of Colour Harmony, *Color Research and Application*, 2010, **35**(1): 34-49, 10.1002/col.20558, ISSN: 0361-2317, Online ISSN: 1520-6378

[15] Szabó F, Bodrogi P., Schanda J. A Colour Harmony Rendering Index based on new Colour Harmony Formulae, *Lighting Research and Technology* 2009 **41**: 165-182, Online ISSN: 1477-0938 Print ISSN: 1477-1535

[16] Brown, D.J., Neale H., Cobb S.V., Reynolds H, The development and evaluation of the virtual city. *International Journal of Virtual Reality*. (1999), **4**(1):.28-41

[17] Standen P.J., BrownD.J. Virtual reality in the rehabilitation of people with intellectual disabilities: Review *Cyberpsychology and Behaviour,* (2005) 8, 3, pp. 272 - 282

[18] Brown D.J., Shopland S., Battersby S., Tully A., Richardson S. (2009) Game On: Accessible serious games for offenders and those at risk of offending. *Journal of Assistive Technologies,* (2009) **3**(2):15-30**.** © Pavilion Journals (Brighton) Ltd.

[19] Sik Lanyi C., Brown D., Standen P, Lewis J.,›Butkute V. (2012) Results of user interface evaluation of serious games for students with intellectual disability, *Acta Polytechnica Hungarica*, 2012, **9(1)**:225-245, ISSN: 1785-8860, http://www.uni-obuda.hu/journal/Issue33.htm

[20] Komlodi, A., Hercegfi, K. Exploring Cultural Differences in Information Behavior Applying Psychophysiological Methods. CHI2010 (ACM Conference on Human Factors in Computing Systems), April 10-15, 2010, Atlanta, GA, USA, Proceedings pp. 4153-4158, ACM Press, ISBN:978-1-60558-930-5: http://portal.acm.org/citation.cfm?doid=1753846.1754118

[21] Clemmensen, T., Hertzum, M., Hornbaek, K., Qingxin, S., Yammiyavar, P. Cultural cognition in usability evaluation. Interacting with Computers (2009) **21**: 212-220

[22] Hall, E.T.The Hidden Dimension. Doubleday, New York, NY, USA, 1990

[23] Gao X-P., Xin J.H., Sato T., Hansuebsai A., Scalzo M., Kaiwara K., Guan S-S., Valldeperas J., Lis M.J., Billger M., Analysis of Cross-Cultural Color Emotion, Color Research & Application, 2007;32(**3**):223-229

[24] Sato T., Kajiwara K., Hoshino H., Nakamura T., Quantitive evaluation and categorization of human emotion induced by colour. *Adv Colour Sci Technol*. 2000;**3**:53-59

[25] Bowman D.A., Kruijff E., Laviola Jr. J.J., Poupyrev I., 3D User Interfaces. Addison-Wesley, 2004

[26] Komlódi A., Józsa E., Hercegfi K., Kucsora Sz., Borics, D. Empirical Usability Evaluation of the Wii Controller as an Input Device for the VirCA Immersive Virtual Space. CogInfoCom 2011, Budapest, Hungary, 2011 July 7-9., Proc. pp.1-6. ISBN 978-1-4577-1806-9

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5999481&tag=1

[27] Komlódi A., Hercegfi K., Józsa E., Köles, M. Human-information interaction in 3d immersive virtual environments. CogInfoCom 2012 - 3rd IEEE International Conference on Cognitive Infocommunications, Kosice, Slovakia, 2012 Dec 2-5. Proc. pp.597-600. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6422049

[28] Steinkuehler, C., Massively Multiplayer Online Games – Based Learning, M3 – Interdisciplinary Aspects on Digital Media & Education, Workshop, Wien, Austria, Nov. 23, 2006, pp. 15-16

[29] Barfield W., Hendrix C., Bjorneseth O., Kaczmarek K.A., Lotens W., Comparison of human sensory capabilities with technical specifications of virtual environment equipments, Presense: Teleoperators and Virtual Environments, **4**, 329-356

[30] Witmer B.G., Bailey J.H., Knerr B.W.. 1996, Virtual spaces and real world places: transfer of route knowledge, International Journal of Human-Computer Studies, **45**, 413-428

[31] Wilson J.R., Virtual environments and ergonomics: needs and opportunities, *Ergonomics*, 1997, **40**(10): 1057-1077

[32] Sik Lányi C., Schanda J.: Analysing the Colours of the Virtual Reality Museum's Picture, Acta Polytechnica Hungarica, 2011, **8(5)**:137-150

[33] MacDonald L. (1998), Color in Computer Graphics, Lecture Notes, 2nd edition, MacColor Ltd.

[34] CIE Draft standard: Colorimetry – Part 4, (2006), CIE 1976 L*a*b* color space, CIE DS 014-4.1

[35] Sik Lányi C., Investigating of Memory – Colors of Intellectually Disabled Children and Virtual Game Addict Students, *Lecture Notes in Computer Science, LNCS 5889*, USAB 2009, Springer Verlag Berlin-Heidelberg, pp. 463-475

[36] Bartleson, C.J., Memory colors of familiar objects. J. OSA 50 73-77. (1960)

[37] Tarczali, T., Investigation of color memory, PhD thesis, University of Pannonia, 2007

[38] CIE TC 1-33 Color Rendering. Specifying Color Rendering Properties of Light Sources, 1996

[39] Sik Lányi C., Sik A., Sik G., What Kinds of Colors are Used in the Virtual Games, 12th International Conference, HCI International 2007, Beijing, China, July 22-27, 2007. Lecture Notes in Computer Science, LNCS 4550-4566, poster, pp. 1285-1288

[40] Sik Lányi C., Végh, V., Schanda, J., Cultural differences in cartoon colours, COST N529 Efficient Lighting for the 21st Century, Workshop WG4 Colour Aspects for Light Sources – Colorimetry and it's Applications in Industry ad Environment, Varna, Bulgaria, pp. 29-31 May 2006, 35-40

Positioning of Public Service Systems Using Uncertain Data Clustering

Ivica Lukić, Mirko Köhler, Ninoslav Slavek

Faculty of Electrical Engineering, Josip Juraj Strossmayer University of Osijek, Cara Hadrijana bb, 31000 Osijek, Croatia
ivica.lukic@etfos.hr, mirko.kohler@etfos.hr, ninoslav.slavek@etfos.hr

Abstract: Positioning of public service system is crucial and very challenging task. Proper positioning ensures that the public service system would complete its tasks to the end users. This paper is focused on finding the best location for public service system, to improve its efficiency when using uncertain data clustering. By choosing the best location for the service system the respond time can be minimised, and the given tasks could be performed in a reasonable time. Improved bisector pruning method was proposed for clustering previous data of public service system to find the best location for its application. Presented method can be used for different Public Service Systems, like traffic services, positioning of ambulance vehicles and other mobile objects. Cluster centres are used as best locations for public service systems because, cluster centres minimized total expected distance from tasks that have been set to the service system. On this way, public service system will be improved and can fulfil more tasks during the shortest period of time.

Keywords: clustering; data mining; expected distance; service systems; uncertain data

1 Introduction

Public Service System (PSS) data are saved in databases and this data can contain uncertainty and, therefore mining useful data from such uncertain database is not a simple task [1, 2]. Different factors, as measurement errors, sampling discrepancy, outdated data source contribute to data uncertainty. To cluster data with location uncertainty deploys various methods such as improved bisector pruning [3], MinMax pruning [4] and Voronoi pruning [5] that were used. Clustered objects are mutually similar, near to the cluster centre and similar objects are positioned in the same group. Cluster centres have minimized the total expected distance from all observed objects. Thus PSS should be located in the cluster centres or near them. Clustering methods are used for tracking of moving objects [6, 7], such as mobile devices, traffic services, ambulance vehicles, etc. Proper positioning ensures that public service system is as close as possible to accomplish its tasks and serve the end users. By choosing the best location, service system would

minimize the distance to accomplish the previously set tasks and also their time to react properly. Data source can be outdated or contain errors and thus object location can contain uncertainty. Uncertain object location is not represented as a discrete point, but as an uncertainty region. In practice, uncertainty region is represented by a Probability Density Function (PDF). In this paper, PSS object's locations were presented in 2D dimensions and a two-dimensional uncertainty. In real life PDF applications can be specified using Gaussian distributions with the means and variances [6]. For Gaussian distributions, density function is exponentially dropped, meaning that probability density outside certain region equals zero. Thus, each object can be bounded by a finite bounding region. This region is limited by the maximum speed of the object and elapsed screening time. Clustering process must have short execution time because efficiency is very important for the PSS applications. Most of the computational time was lost on expected distances (ED) calculations [8, 9]. Large number of sample points that were used to represent each PDF [10], and a numerical integration is involved to calculate expected distance. Distance had to be calculated for all samples, thus computational cost were higher than in a simple distance calculation [11]. In [4] and [5], the pruning methods are introduced and thus the ED calculations can be avoided. In this study, the bounding regions may represent the pruning objects. Using these pruning methods, some clusters are eliminated as candidate clusters, if the closer cluster for observed object had been found. In [3] the improved bisector pruning method is presented to improve the clustering process. It is compared to the existing methods and it was experimentally proved that it had the best clustering results and shortest execution times in most of the situations.

2 Improved Bisector Pruning

Uncertain objects are data collection $O=\{o_1,...o_n\}$ in m dimensional space R^m. Distance between two objects is always greater than zero:

$$d(o_i, o_j) \geq 0 \tag{1}$$

Probability density function of each object at each point $x \in R^m$ is $f_i(x) > 0 \quad \forall\, x \in R^m$, where for all points inside MBR is:

$$\int_{x \in R^m} f_i(x)dx = 1 \tag{2}$$

Expected distance from object o_i to any point y is calculated using the formula:

$$ED(o_i, y) = \int_{x \in A_i} d(x, y) f_i(x)dx \tag{3}$$

Bounded region A_i is finite region and $f_i(x)=0$ is set for outside that region. Goal of clustering is to find set of clusters points $C=\{c_1,...c_m\}$ and all relations between objects and clusters $h:\{1,...,n\}\rightarrow\{1,...,m\}$, for which total expected distance (TED) from all the objects assigned to cluster centre is minimised, as shown in formula (4). $TED = \sum_{i=1}^{n} ED(o_i, c_{h(i)})$ (4)

Improved Bisector Pruning inherits principles of Voronoi and Bisector pruning method and it is improvement of the aforementioned method [5, 12]. Bisector pruning is a side product of Voronoi diagrams construction, and bisectors are calculated as a small additional calculation cost after Voronoi diagrams was constructed. Thus, Bisector pruning is combined with the Voronoi cell pruning [12]. In Improved Bisector pruning, Voronoi diagrams are not constructed and bisectors are calculated using formula (7), which is more effective than Voronoi pruning. Bisector is a line segment that is perpendicular to the line segment joining c_p and c_q, and that passes through the mid-point of the line segment. For each pair of clusters c_p and c_q in $C=\{c_1,...c_m\}$ bisector $B_{p/q}$ is calculated using the following formulas:

$$a = -\left(\frac{x_{cp} - x_{cq}}{y_{cp} - y_{cq}} \right) \tag{5}$$

$$b = \frac{x_{cp}^2 - x_{cq}^2 + y_{cp}^2 - y_{cq}^2}{2(y_{cp} - y_{cq})} \tag{6}$$

$$B_{p/q} = a * x + b \tag{7}$$

All bisectors are constructed using representative cluster points (x_{cp}, y_{cp}) and (x_{cq}, y_{cq}). For each cluster pair it was checked that if MBR_i of object o_i completely lies on the same side of bisector $B_{p/q}$ as cluster c_p, and if it does so, cluster c_q is pruned from object o_i. For the opposite situation, if MBR_i of object o_i completely lies on the same side of bisector $B_{p/q}$ as a cluster c_q, then cluster c_p is pruned from object o_i. Pruned cluster is instantly removed from cluster candidates. For 50 clusters there are 50 x 50 bisectors calculations. But for once pruned cluster, remaining bisectors are not constructed and this number is significantly reduced. For the clusters to be pruned, next properties must be satisfied:

$$\begin{aligned} &(y_{bcp} > y_{cp} \ and \ y_{boi} > y_{oi}) \\ &or \ (y_{bcp} < y_{cp} \ and \ y_{boi} < y_{oi}) \end{aligned} \tag{8}$$

In Figure 1 are explained principles used in above formula. For the points on the perimeter of MBR_2 it is checked do they lie on the same side of bisector as cluster c_3. To check this statement coordinate x_{c3} of cluster c_3 and coordinate x_{o2} of point on the perimeter of MBR_2 are included in bisector formula, and results y_{bc3} and y_{bo2} are obtained.

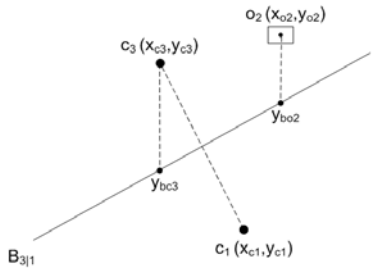


Figure 1

An example of Bisector pruning where projected coordinates are lower than original coordinates

From Figure 1 it is obvious, that object o_2 and cluster c_3 are on the same side of the bisector $B_{3/1}$ according to the formula (8). Obtained result y_{bc3} is lower than original coordinate y_{c3} of cluster c_3, and also y_{bo2} is lower than original coordinate y_{o2} of point on perimeter of MBR_2.

In the opposite situation, as it was shown in Figure 2, obtained result y_{bc3} is higher than y_{c3}, and y_{bo2} is higher than y_{o2}. Condition that was set in formula (8) is satisfied and object o_2 is on the same side of the bisector as cluster c_3. All the aforementioned steps must be repeated for the peak points on MBR_2. If all points satisfy the formula (8), cluster c_1 is pruned from object o_2.

After iterating all cluster pairs, we will find that the most of the clusters were pruned, and only for few remaining clusters ED calculation will be needed. In Voronoi pruning, if all clusters except one cluster are not pruned, ED must be calculated for all the clusters. Thus, Voronoi diagram is combined with Bisector pruning, to prune remaining clusters and thus avoid all the

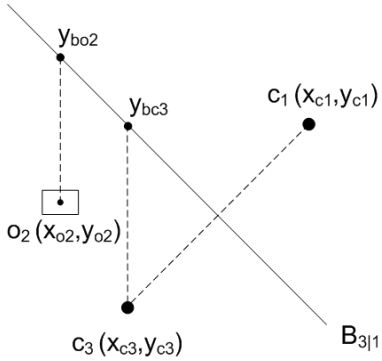


Figure 2

Example of Bisector pruning where projected coordinates are higher than original coordinates

unnecessary ED calculations. Besides, calculations that are using formulas (1-5) are faster than Voronoi diagrams construction, and for the each object must be

checked does its MBR_i completely lies inside the Voronoi cell. Improved Bisector pruning is described by the following algorithm:

for all distinct $c_p, c_q \in C$ **do**

if MBR_i on the same side of bisector $B_{p/q}$ as cluster c_p **then**

Remove c_q **from** CC_i */*candidate clusters*/* Remove c_q **from** iteration loop **(for)**

else if MBR_i on the same side of bisector $B_{p/q}$ as cluster c_q **then**

Remove c_p **from** CC_i */*candidate clusters*/*

Remove c_p **from** iteration loop **(for)**

for all the remaining candidate clusters calculated ED

2.1 Combination with the SDSA Method

Improved Bisector pruning method is combined with SDSA (Segmentation of Data Set Area) to shorten execution time of clustering process [13]. In SDSA method data set area was divided into small segments. Segments are parts of a total data set area, as it was shown in Figure 3. All segments have the same size, and they are rectangular. Observed were only objects in one segment and they pairs with clusters from that and neighbouring segments. Thus, number of object clusters observations was decreased. Experiments have showed that SDSA method combined with other pruning methods speeds up clustering process, depending on the number of clusters and objects [13]. Improvement of clustering process is reverse proportional to the size of segments. If segments are smaller than the clustering process is more effective. By decreasing the size of segments, the number of observed object clusters pairs is decreased, as shown in Figure 3.

An entire data set area C is shown in Figure 3a, and in Figure 3b data set area was divided into four small clusters segments SSDSA. Each segment was divided into four smaller segments. In Figure 3c are showed 16 small segments SSDSA and that was the final number of segments observed in this paper. Enlarged areas with object set inside them are shown in the Figure 3d, and Figure 3e. For 1600 objects and 64 clusters, there are 102400 object cluster pair calculations. If the SDSA method were used, the data set area was divided into 16 smaller segments of SSDSA, and clusters area C was divided into 4 small areas CSDSA. The average number of objects in one segment is 100, and the average number of clusters is four. Each segment was observed separately. Total number of observed objects was 100, number of observed clusters was 16 and the number of segments was also 16. Thus, total number of object cluster pair calculations was 25600, what

represented the four times less calculations to proceed with. In this case, there was no need to check for all the clusters, but only clusters which are near one another and those who surrounded the area. All remaining clusters are pruned for all the objects that were contained inside the area. In this case, SDSA pruning decreases the total number of calculations by four times. When decreasing of the total number of calculations was done, it was proportional to decrease of the areas with clusters. However, segmentation had size limits, which are dependent on the number of clusters and their positioning. Segments have to be surrounded by clusters, and if the number of clusters is high, then segments can be very small and speed up the clustering process.

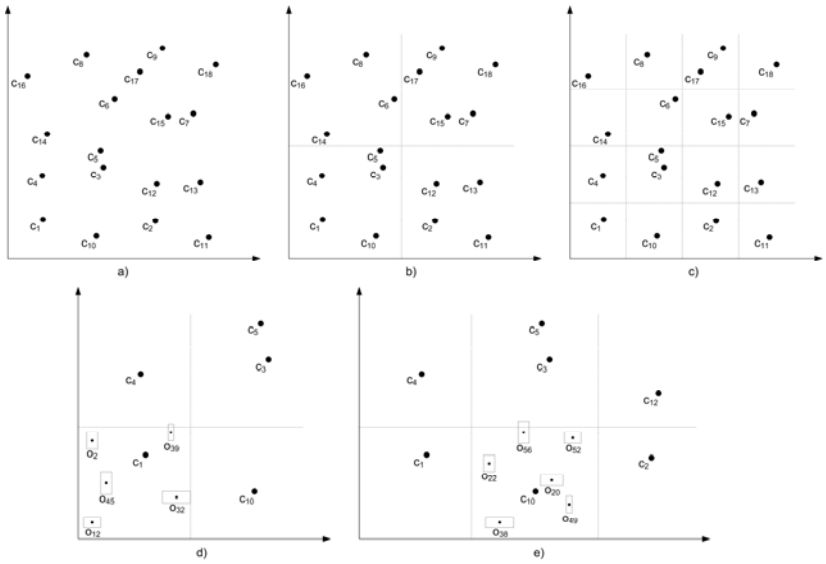


Figure 3
Data set area process of segmentation

3 Experiments

For clustering the existing PSS data, improved Bisector pruning method was used. In these experiments, as an emergency response public service vehicles in the city of Osijek were used. During the previous two months all the data were collected and processed to find the best location for positioning of the vehicles. Without clustering, vehicles were located in one place and needed more time to arrive on the remote locations. Arrival time is critical in some situations and human lives can depend on it. By clustering the existing data about PSS we could find cluster centres as the best locations that minimize total distance from executing possible

tasks. Vehicles should be located in cluster centres and wait to be called upon for the fulfilment of the next task. All the new tasks were assigned to the waiting vehicles in the nearest cluster centre. It was found that vehicles from the cluster centre could accomplish the set task much faster than vehicle located in remote peripheral areas. In this paper, experiments were conducted with aims to prove that clustering existing service system data improves reaction time of PSS. In the set experiment, distance from cluster centres to tasks that have to been accomplished was measured and compared to the distance from existing central places to the set tasks. All experiments data were implemented in MATLAB 7.0 and carried out on a PC with an Intel Core(TM) 2 Duo Mobile CPU at 2.00 GHz, and 1.75 GB of main memory (RAM) All locations were presented in geo – coordinates system, where φ represents latitude, λ longitude and *r* radius of the Earth. Clustering methods are designed for clustering the Cartesian coordinates and conversion from spherical coordinates that had to be made. From Figure 4 is visible that conversion from spherical to Cartesian coordinates can be done using the following formulas:

$$x = r\cos\varphi\cos\lambda \tag{8}$$

$$y = r\cos\varphi\sin\lambda \tag{9}$$

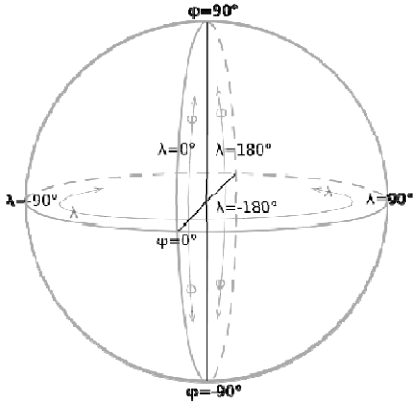


Figure 4
Spherical geo - coordinates representation.

3.1 Experiments Done When Using Three Clusters

In this experiment, all existing data about emergency responses in the city of Osijek were clustered in total of three clusters. Depending on the number of tasks that needed to be accomplished and which surround the cluster area, in each cluster centre should be located the exact number of vehicles. Three cluster centres are presented as red dots in Figure 5.

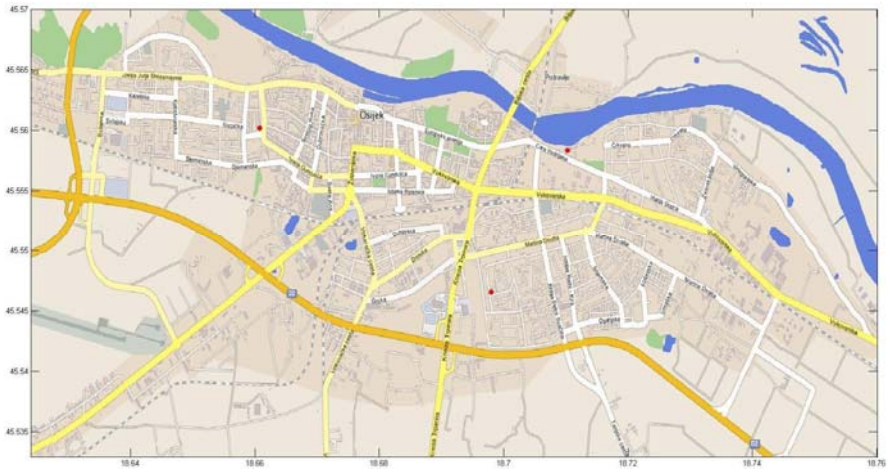


Figure 5
Public service system consisting of three cluster centres

In Table 1 is presented the number of tasks for each cluster centre. The most tasks are located in cluster centre No. 3, and therefore emergency response system should locate more vehicles in the surrounding area. Fewer vehicles were located in the remaining two cluster centres, according to the number of tasks for each cluster centre.

Table 1
Number of tasks settled in each of the cluster centres

Cluster centre	*Number of tasks*
Centre No. 1	362
Centre No. 2	531
Centre No. 3	744

Efficiency and reaction time of the vehicles were improved by positioning the vehicles effectively, when considering the PSS. When positioning the vehicles in the cluster centres, it can be ensured that total foreseeable distance from PSS that connects its tasks and users who use the service, is minimised. Total expected distance is reduced up to 40%, when compared to the situation without clustering been considered.

3.2 Experiments Done When Using Four Clusters

In this experiment all existing data about emergency responses in the city of Osijek were clustered in four clusters. Depending on the number of tasks that needed to be accomplished, and which surround the cluster area, in each cluster centre should be located the exact number of vehicles. Cluster centres are showed as red dots in Figure 6.

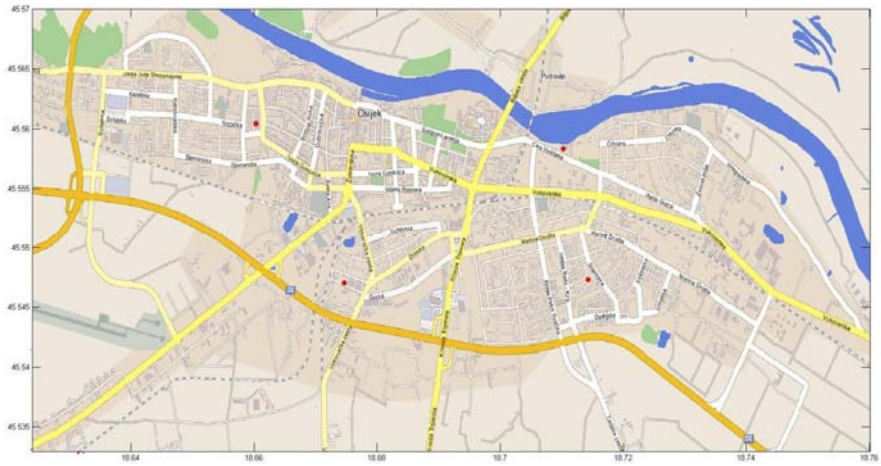


Figure 6
Public service system consisting of four cluster centres

In Table 2 is presented the number of tasks for each cluster centre. The most tasks are located in cluster centre No. 4, and therefore emergency response system should locate more vehicles in the surrounding area. Fewer vehicles were located in the remaining cluster centres, according to the number of tasks for each cluster centre. Total expected distance is reduced up to 40%, when compared to the situation without clustering been considered.

Table 2
Number of tasks settled in each of the cluster centres

Cluster centre	*Number of tasks*
Centre No. 1	295
Centre No. 2	317
Centre No. 3	334
Centre No. 4	691

3.3 Experiments Done When Using Five Clusters

In this experiment all existing data about emergency responses in the city of Osijek were clustered in five clusters. Depending on the number of tasks that needed to be accomplished, and which surround the cluster area, in each cluster centre should be located the exact number of vehicles. Cluster centres are showed as red dots in Figure 7.

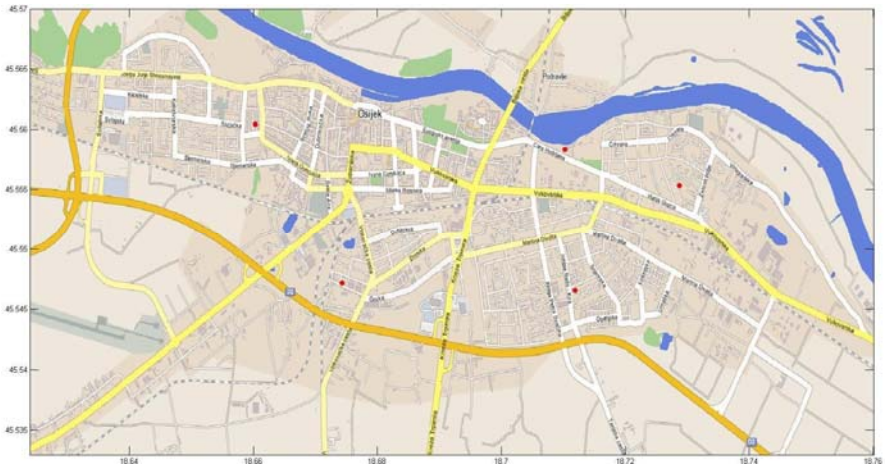


Figure 7
Public service system consisting of five cluster centres

In Table 3 is presented the number of tasks for each cluster centre. The most tasks are located in cluster centres Nos. 2 and 5 and therefore, emergency response system should locate more vehicles in the surrounding area. Fewer vehicles were located in the remaining cluster centres, according to the number of tasks for each cluster centre. Total expected distance is reduced up to 50%, when compared to the situation without clustering been considered.

Table 3
Number of tasks settled in each of the cluster centres

Cluster centre	*Number of tasks*
Centre No.1	227
Centre No. 2	482
Centre No. 3	375
Centre No. 4	86
Centre No. 5	467

4 Dedication to the Future Work

During our future work, we will create a database for storage of the geographical coordinates of all addresses in Osijek. Streets were divided into sections that represent one uncertain object with a minimum bounding region. Uncertain object is sampled, and to each sample is assigned probability for that object located in that very sample. The number of samples is proportional to the number of addresses in one city block, because every address is represented by one sample.

The probability that the object is located in certain sample represents a number of tasks that should be performed by the public service system when that observed sample is being considered. Database will contain all the tasks that service system should perform in each sample. For each pattern, the number of tasks that have to be done can be found in the database, and the value of the probability density function can be calculated. Based on the current data from the database we can predict future tasks of the public service system. The aforementioned prediction can significantly reduce the respond time and save funds for the public service system maintenance. By performing the public service data system, we can determine cluster centres that represent regions with the greatest concentration of tasks, for the operations of the public service system. Using predictive models, we can enable the public service system to deploy their workers in cluster centres and to efficiently accomplish the needs of their customers. Tasks of public service system are dependent on various unexpected events that could change the requests upon them. In our future work we will conduct experiments to predict tasks for the most unexpected events. For the purposes of the experiments we will make the script that shall contain the geographic coordinates of all addresses in Osijek. Experiments and simulations will be conducted for the normal operation of public service systems, and also experiments for the unexpected events where the number and site locations of tasks were significantly changed. Using predictive and existing data about similar events in the past, public service system will know where tasks might happened and properly deploy their employees to the target locations. The proposed model will reduce the costs of the public service system, enhanced the number of tasks that the can be done in a given period of time, and will also increase the reliability of the public service system.

Conclusion

This paper discussed an important role for choosing the best location for public service system. For this purpose, uncertain data clustering methods were used. Improved Bisector pruning method was used for clustering previous aggregated data of public service system to find the best location for their operational service. Cluster centres are used as best location for public service systems, because such centres minimized the total expected distance for the tasks that had been set to fulfil the service system operations. Several experiments were conducted. In the first experiment, public service system was divided into three clusters; in the second experiment in four clusters and finally, in the last experiment, in five clusters. Experiment with four clusters was used for calculation of distance reduction, because clusters in this experiment are nearest to the existing public service system locations set. It was experimentally proved that positioning of public service system can improve effectiveness and reaction time of such system, when comparisons were drawn to the existing system. Proper positioning ensures that public service system is as close as possible to accomplish its tasks and serve the end users. By choosing the best location, service system would minimize the distance to accomplish the previously set tasks and also their time to react

properly. Our analyses have shown that the operating distances were reduced up to 50%. On this way, public service system will improve and can accomplish more tasks during the same period of time. In our future work, we will concentrate to further improve our model with additional parameters. We will try to apply our model to the other public service systems to improve existing systems and reduce its costs and also the service response time.

Acknowledgements

This work was supported by research project grant No. 165-1652017-2016, issued by the Ministry of Science, Education and Sports of the Republic of Croatia.

References

[1] D. Nilesh and D. Suciu, *"Efficient query evaluation on probabilistic databases"*. In Proc. of VLDB Conference, pages 864–875, 2004

[2] R. Cheng, X. Xia, S. Prabhakar, R. Shah, and J. Vitter, *"Efficient indexing methods for probabilistic threshold queries over uncertain data"*. In Proc. of VLDB Conference, 2004

[3] Lukić, M. Köhler, N. Slavek, "*Improved Bisector Pruning for Uncertain Data Mining*", Proceedings of the 34th International Conference on Information Technology Interfaces, ITI 2012., pages 355-360

[4] W. K. Ngai, B. Kao, C. K. Chui, R. Cheng, et al. "Efficient clustering of uncertain data". In ICDM, pages 436–445, 2006

[5] B. Kao, S. D. Lee, D. W. Cheung, W. S. Ho, K. F. Chan, *"Clustering Uncertain Data using Voronoi Diagrams"*. Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on Date:15-19 Dec. 2008. On pages: 333 – 342

[6] O. Wolfson, P. Sistla, S. Chamberlain, and Y. Yesha, *"Updating and querying databases that track mobile units"*. Distributed and Parallel Databases, 7(3), 1999.

[7] R. Cheng, D. Kalashnikov, and S. Prabhakar, *"Querying imprecise data in moving object environments"*. IEEE TKDE, 16(9):1112–1127, 2004.

[8] J. MacQueen, *"Some methods for classification and analysis of multivariate observations"*. In Proc. 5th Berkeley Symposium on Math. Stat. and Prob., pages 281–297, 1967.

[9] M. Ichino and H. Yaguchi, *"Generalized minkowski metrics for mixed feature type data analysis"*. IEEE TSMC, 24(4):698V–708, 1994.

[10] L. Xiao, E. Hung, *"An Efficient Distance Calculation Method for Uncertain Objects"*. Computational Intelligence and Data Mining, CIDM 2007., pages 10–17, 2007.

[11] M. Chau, R. Cheng, B. Kao, and J. Ng, *"Uncertain data mining: An example in clustering location data"*. In PAKDD, pages 199–204, Singapore, 9–12 Apr. 2006. Springer.

[12] B. Kao, S. D. Lee, F.K.F. Lee, D. W. Cheung, W. S. Ho, *"Clustering Uncertain Data using Voronoi Diagrams and R-Tree Index"*. Knowledge and Data Engineering, IEEE Transactions, Sept. 2010. On pages: 1219-1233

[13] Lukić, M. Köhler, N. Slavek, "The Segmentation of Data Set Area Method in clustering of Uncertain Data", Proceedings of the jubilee 35th International ICT Convention – MIPRO 2012., pages 420-425

Evaluation of Flexible Graphical User Interface for Intuitive Human Robot Interactions

Balázs Dániel[1], Péter Korondi[1], Gábor Sziebig[2], Trygve Thomessen[3]

[1] Department of Mechatronics, Optics, and Mechanical Engineering Informatics, Budapest University of Technology and Economics, PO BOX 91, 1521 Budapest, Hungary
e-mail: daniel@mogi.bme.hu; korondi@mogi.bme.hu

[2] Department of Industrial Engineering, Narvik University College, Lodve Langes gate 2, 8514 Narvik, Norway
e-mail: gabor.sziebig@hin.no

[3] PPM AS, Leirfossveien 27, 7038 Trondheim, Norway
e-mail: trygve.thomessen@ppm.no

Abstract: A new approach for industrial robot user interfaces is necessary due to the fact that small and medium sized enterprises are more interested in automation. The increasing number of robot applications in small volume production requires new techniques to ease the use of these sophisticated systems. In this paper shop floor operation is in the focus. A Flexible Graphical User Interface is presented which is based on cognitive infocommunication (CogInfoCom) and implements the Service Oriented Robot Operation concept. The definition of CogInfoCom icons is extended by the introduction of identification, interaction and feedback roles. The user interface is evaluated with experiments. Results show that a significant reduction in task execution time and a lower number of required interactions is achieved because of the intuitiveness of the system with human centered design.

Keywords: industrial robotics; human robot interaction; cognitive infocommunication; flexible robot cell; graphical user interface; usability

1 Introduction

In the era of touch screen based smartphones the interaction between humans and machines is becoming a frequent event. Especially the communication between people and robots is a field of research nowadays. Human Robot Interaction (HRI) studies are aiming the user friendly application of different robot systems for

cooperation with humans. This includes an interactive conversation robot which collects multi-modal data and bonds with human partners. [1]

While the interaction with an industrial robot is traditionally considered as a Human Machine Interaction (HMI) due to its inferior level of autonomy and complexity [2], these systems are also spreading to human environments. Shared workspace [3], teach by demonstration [4], all these improvements in industrial robotics require a higher level of communication. The need for better and more intuitive user interfaces for industrial manipulators is increasing.

The reason of the increasing interest is that a wide range of Small and Medium Sized Enterprises (SMEs) are motivated to invest in industrial robot systems and automation due to increasing cost levels in western countries. The last years a significant part of robot installations in industry required high flexibility and high accuracy for low-volume production of SMEs [5]. This tendency adds new challenges to support and transfer knowledge between robotics professionals and SMEs. The difficulty of operating a robotic manipulator is present both on shop-floor production and remote support for the robot cell. Operators with less expertise will supervise the robot cells in SMEs because of economical reasons and there is usually no dedicated robot system maintenance group for small companies. Remote operation of the industrial robot systems from the system integrator's office offers an alternative solution. This is based on the premise that if one can perform remote assistance for the SMEs, the enterprise can get high reliability, high productivity, faster and cheaper support irrespective of geographical distance between the enterprise and the system integrator or other supportive services. On the other hand the need for assistance from professionals can be reduced by introducing such user/operator interfaces which fit better for inexperienced users and provide an intuitive and flexible manner of operation.

The paper is organized as follows: Section 2 briefly summarizes the properties of Human Robot Interaction for industrial robots. Section 3 presents a Flexible Graphical User Interface implementation test with quantitative and qualitative results.

2 Human Robot Interaction

Scholtz [2] proposed five different roles for humans in HRI. Supervisor, operator, mechanic, teammate and bystander; these define the necessary level of information exchange from both sides. In order to keep the interaction continuous the switch between roles is inevitable in certain situations. Efficient user interfaces have to take into account the adequate display of details in complex systems like robot cells.

Automated systems may be designed in machine-centered or human-centered way [6]. While the first technique requires high level knowledge from the user for operation, the latter is more adaptable for flexible systems which serves both seasoned and unseasoned personnel. Standard HRI for industrial equipment is usually utilize machine-centered approach whereas complex systems are accessible through complicated user interfaces.

Human-centered design applies multidisciplinary knowledge, thus combining the necessary factors and benefits both for machinery and human operation. The ISO Standard 9241-210:2010 [7] defines it as an activity to increase productivity in parallel with improved work conditions by the use of ergonomics and human factors. User friendly or human friendly systems are designed with the aim of minimizing the machine factor. This implies that in this case robots have to adapt to the operator, but the difficulty is to provide a comprehensive and appropriate human behavior model [8].

Human factors include psychological aspects also. HRI should be analyzed from this point of view to make a better match between robot technology and humans. The systematic study of the operator's needs and preferences is inevitable [9] to meet the expectations on psychological level, moreover, in order to satisfy these needs, multimodal user interfaces are necessary which may facilitate the communication with intuitive and cognitive functions [9, 10].

Figure 1
Modern Teach Pendants.
From top-left corner: FANUC iPendant [11], Yaskawa Motoman NX100 teach pendant [12], ABB FlexPendant [13], KUKA smartPad [14], ReisPAD [15], Nachi FD11 teach pendant

For industrial robotic manipulators the standard user interface is the Teach Pendant. It is a mobile, hand-held device usually with customized keyboard and graphical display. Most robot manufacturers are developing distinctive design (See Figure 1) and are including a great number of features thus increasing the complexity and flexibility.

Input methods vary from key-centric to solely touch screen based. Feedback and system information is presented on graphical display in all cases, however the operator may utilize a great number of other information channels; touch, vision, hearing, smell and taste. The operator can also take an advantage of the brain's ability to integrate the information acquired from his or her senses. Thus, although the operator has the main sight and attention oriented towards the robot's tool, she will immediately change her attention to any of the robot's link colliding with an obstacle when it is intercepted by her peripheral vision.

The science of integrating informatics, info-communication and cognitive-sciences, called CogInfoCom is investigating the advantages of combining information modalities and has working examples in virtual environments, robotics and telemanipulation [10, 16]. The precise definition and description of CogInfoCom is presented in [17], terminology is provided in [18], while [19] walks the reader through the history and evolution of CogInfoCom.

Introducing this concept in the design of industrial robot user interfaces is a powerful tool to meet the goal of human-centred systems and a more efficient human robot interaction.

3 Flexible Graphical User Interface

At first the focus is laid on the possible improvements regarding the existing Teach Pendant information display. The traditional industrial graphical user interfaces are designed to deal with a large number of features, thus the organization of menus and the use of the system is complex and complicated.

In a real life example the robot cell operator had to ask assistance from the system integrator in a palletizing application because the manipulator constantly moved to wrong positions. Without clear indication on the robot's display the user could not recognize that the programmed positions are shifted due to the fact that the system is in the middle of the palletizing sequence. The resetting option was in two menu-levels deep in the robot controller's constant settings. In case of a flexible user interface this could have been avoided by providing custom surface for palletizing instead of showing the robot program, the date, etc.

In most cases the customization of the robot user interface is possible. A separate computer and software is required [20] and due to the complexity of the system it

is done by the system integrator. Naturally, the integrator cannot be fully aware of the needs of the operator thus the flexibility of a robot cell is depending also on the capabilities of the framework in which the graphical user interface is working. The communication barrier between the operator and the integrator influences the efficiency of the production through inadequate and non-optimized information flow (See Figure 2). Differences in competence level, geographical distance or cultural distance can all cause difficulties in overcoming this barrier.

The Flexible Graphical User Interface (FGUI) concept aims to close this gap. The system integrator still have the opportunity to compile task specific user interfaces but the framework includes pre-defined, robot specific elements which are at the operators' hands at all times. The functionality is programmed by the integrator and represented as she thinks it is the most fitting, but the final information channel can be rearranged by the user/operator to achieve efficient and human (operator) centered surface.

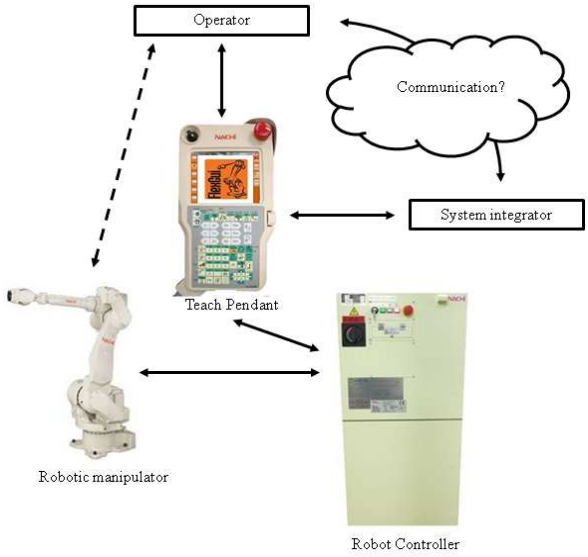


Figure 2
Connection of a robot system with the operator and the integrator

3.1 Shop-floor operation

During shop-floor operation the user should communicate on high level with the robot cell. In high volume and highly automated production lines this interaction is restricted the best to start a program at the beginning of the worker's shift and stop it when the shift is over. For SMEs this approach is not feasible: frequent reconfiguration and constant supervision is inevitable. Therefore a service-

oriented user interface is favourable which hides the technical details of a multipurpose robot cell behind an abstract level. This abstraction should be made by utilizing the principles of CogInfoCom and allowing intuitive use of a robot cell.

The connections between the actual robot programs and the operator's user interface are presented in Figure 3. As the system keeps track of the parts in a pick and place operation this data is stored as a set of variables. The operator can observe these values by traditional display items but a FGUI allows to present it in a different way: small lamps are giving visual information of the system's state. The operator instinctively and instantly can compare the actual state of the robot cell with the information stored in the robot controller and detects defects faster. This is an application of representation-bridging in CogInfoCom, because the information on the robot variable is still visual but the representation is different from standard robot variable monitoring. The direct observation of variables would require the knowledge of the system integrator's convention of data representation from the operator.

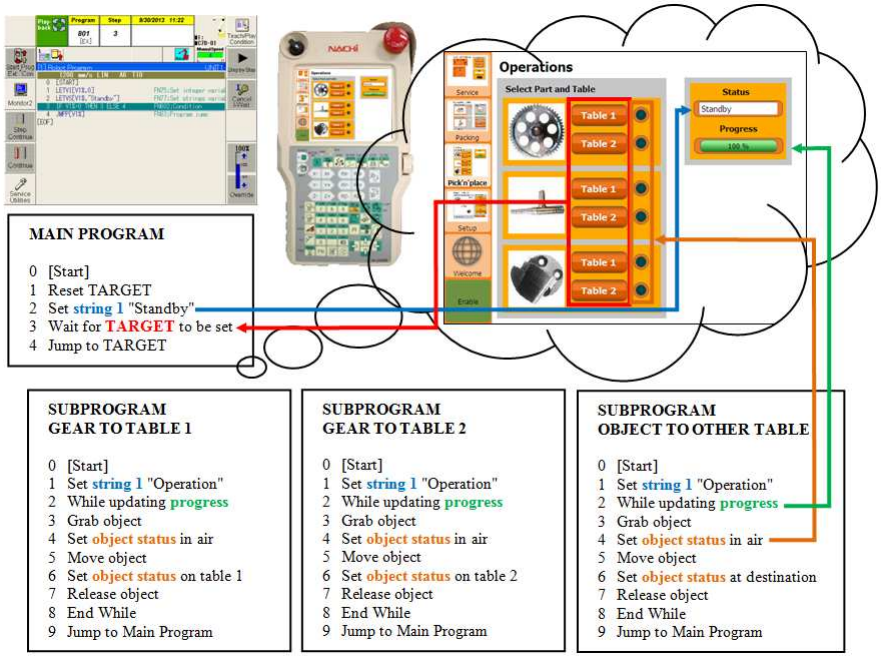


Figure 3

Connections between functionality and service-oriented layers

The selection of programs for the operation is reduced to service requests initiated by button presses paired with an image of the part. Using CogInfoCom semantics [18]; the images of parts, the buttons, the indicator lamps, the progress

bar and the text box are all visual icons bridging the technical robot data with operation parameters and statuses. The high-level message generated by these entities is the list of achievable item movements in case of a pick and place service and feedback on the current operation provided by the user interface not from visual observation of the cell thus we consider this application of CogInfoCom channel as an eyecon message transferring system in general.

Furthermore, considering a group of screen elements (e.g. the image of the gear, the buttons "Table 1" and "Table 2" and the indicator lamps) we use low-level direct conceptual mapping. The picture of the part generates an instant identification of the robot service objective and the surface to interact with the robot controller. The buttons represent identification and also interaction with the robot cell. The robot system internal mapping of real world is sent back to the user through the indicator lamps in feedback form.

These applications indicates that CogInfoCom icons usually not only generate and represent a message in the communication but these also have roles. These roles may depend on the actual implementation and concept transmitted by messages. In human robot interaction three main roles may be distinguished:

- identification role,
- interaction role,
- feedback role.

The simple instruction "Move the gear from Table 1 to Table 2" given to the operator does not require additional mapping from the user thus the human robot interaction is simplified and became human centered. The gear may be identified by image, manipulated by interaction with the button and get feedback though the indicator lamp.

Practically the user interface implements an abstract level between the technical realization of operations and the service oriented operation. A main robot program is monitoring the inputs from the user. By pressing one of the buttons a number is loaded into a variable which causes to controller to change to the program indicated by this number. The program contains the pre-defined pose values to be executed sequentially. The current program line number divided by the total number of lines indicates the progress with task for the user in the progress bar, since one program contains only one item movement from one table to another. Exiting the main program also sets the text box to "Operation in progress" and re-entering it resets to "Standby".

The robot controller keeps track of objects in the robot cell by adjusting an integer variable according to the current position of it in the motion sequence. Values 1, 2 and 0 represent the item placed on Table 1, placed on Table 2, and lifted and moved by the robot, respectively. Mapping this information onto the indicator lamps means that when the object is on one of the tables, the lamp next to the

table button lights up, sending the message that this table is occupied. When it is in the air both lamps are dimmed.

It is generally important to rate the success of a concept and its implementations with experiments. The assessment in communication improvement is may be measured both quantitatively and qualitatively.

3.2 Concept evaluation with user tests

Justification of the service-based flexible user interface is performed by a series of testing with real users. It was a comparative test of a traditional and very conservative system and a newly developed flexible graphical user interface. The robot cell was set up for two different services which are the following:

- pick and place operation with two positions on three work pieces,
- parameterizable delivery system for bolting parts.

3.2.1 Test description

Test participants executed two main tasks first using the traditional system (Figure 4) then repeating them using the flexible user interface (Figure 5 and Figure 6). During the tests user interactions with the robot system were recorded by cameras, key logging, and screen capturing.

Figure 4
Traditional Graphical User Interface

At the beginning of each task a briefing was given to the participants on the general aim of the executed task (Table 1). After that they read through a summary on the necessary settings (Table 2 and Table 3) to be made before the robot can perform its operation. At this point participants had the opportunity to go on with a user manual with step-by-step instruction on the necessary steps or continue on their own. Chosen the second option they always had the chance to refer to the manual if they needed suggestions on how to continue.

Figure 5
Flexible Graphical User Interface for Task 3

Figure 6
Flexible Graphical User Interface for Task 4

During the execution of Task 1 the only action needed was to select the appropriate program for the robot based upon the required part movement. Task 2 was more complicated; three internal variables had to be set according to the number of nuts and bolts and the number of the delivery box.

The participant was provided with the new user interface for the execution of Task 3 and Task 4 thus the direct interaction with robot controller variables and setting were obscured.

Table 1
General aims of tasks

Task ID	Description	Interface[1]
Task 1.1 Task 1.2	Move the pipe from Table 1 to Table 2 Move the cutting head from Table 2 to Table 1	TGUI
Task 2	Deliver two bolts and three nuts to Box1	TGUI
Task 3.1 Task 3.2	Move the pipe from Table 2 to Table 1 Move the cutting head from Table 1 to Table 2	FGUI
Task 4	Deliver two bolts and three nuts to Box 1	FGUI

Table 2
Settings for Task 1

Item	Source	Destination	Program Number
Pipe	Table 1	Table 2	816
	Table 2	Table 1	815
Cutting head	Table 1	Table 2	814
	Table 2	Table 1	813
Gear	Table 1	Table 2	812
	Table 2	Table 1	811

Table 3
Settings for Task 2

Parameter	Variable	Value
Program	Program Number	821
Number of nuts	Integer Nr. 014	3
Number of bolts	Integer Nr. 015	2
Delivery box ID	Integer Nr. 016	1

3.2.2 Results

The evaluation was conducted with four participants, all male, between age 25 and 27. All four have engineering background; two have moderate, two have advanced experience in robot programming. Participants were advised that audio and video is recorded which serves only for scientific analysis and they were assured that the test can be interrupted anytime on their initiation.

All participants were able to execute all of the tasks in a reasonable amount of time. Two users reported difficulties to set the program number during Task 1.2. The problem turned out to be a software bug in the traditional robot controller user interface; the user manual had been modified accordingly, although none of the

[1] TGUI: Traditional Graphical User Interface, FGUI: Flexible Graphical User Interface

participants followed the user manual steps strictly, most likely due to their previous experience with robots.

Figure 7
Synchronised videos had been taken during testing

Collected data have been evaluated after the tests. Three parameters have been selected to represent the difference between the traditional and the flexible approach: task execution duration, number of interactions and ratio of number of key presses to touch screen commands.

Execution time is measured between the first interaction with the Teach Pendant recorded by the key logger and mouse logger running on the robot controller, and the last command which ordered the robot to move. The last interaction was determined by time stamp on the video in case of the TGUI measurement, since logging of program start button on the controller housing was not in place at the time of the experiment. The start of robot movement could be determined from mouse logging data in case of FGUI.

The mean execution time using the conservative system for Task 1.1, Task 1.2 and Task 2 are 36, 37 and 135 seconds respectively. In contrast, Task 3.1, Task 3.2 and Task 4 duration was 31, 0 and 30 seconds with the use of the FGUI. Data are presented in Table 4. The total mean time spent on Task 1 and Task 3 are 72,1 and 38,1 second respectively. All the mean values presented based on three measurements. Due to the low number of measurements further statistical data were not computed.

Table 4
Task execution time

	Task duration [s]					
	1.1	1.2	2	3.1	3.2[2]	4
User 1	50,0	53,5	131,5	64,5	0	21,7
User 2	25,2	30,9	121,0	27,9	0	11,1
User 3	31,6	25,1	151,3	21,8	0	22,4
Mean	**35,6**	**36,5**	**134,6**	**38,1**	**0**	**18,4**
User 4[3]	*N/A*	*N/A*	*126,3*	*9,4*	*0*	*63,8*

Participants had to configure the robot controller using the Teach Pendant which offered two possible interactions: key presses and touch screen interactions. A virtual representation of the keyboard is depicted in Figure 7. Besides, users had to deactivate the joint brakes and energize the motors as well as start the robot program by pressing buttons on the robot controller housing. All interactions were counted by the logging software and later the touch interaction ratio to key presses was calculated as follows:

$$TR\ [\%] = \frac{n_{touch}}{n_{touch} + n_{keypress} + n_{controller}} \cdot 100\,, \tag{1}$$

where n_{touch} is the number of touch screen interactions, $n_{keypress}$ is the number of button presses on the Teach Pendant and $n_{controller}$ is the number of button presses on the controller housing. The total number of interactions and touch ratio is listed in Table 5. The data loss in execution time measurement did not affect counting of interactions thus mean values are calculated from four samples.

Table 5
Number of interactions and touch ratio

	Interactions [-] (*Touch ratio [%]*)					
	1.1	1.2	2	3.1	3.2	4
User 1	7 (*14*)	21 (*19*)	57 (*11*)	7 (*71*)	1 (*100*)	10 (*100*)
User 2	7 (*14*)	10 (*30*)	33 (*15*)	6 (*50*)	1 (*100*)	6 (*100*)
User 3	7 (*14*)	14 (*50*)	53 (*15*)	6 (*83*)	1 (*100*)	8 (*100*)
User 4	7 (*14*)	13 (*23*)	26 (*17*)	3 (*67*)	1 (*100*)	7 (*71*)
Mean	**7 (*14*)**	**15 (*31*)**	**42 (*15*)**	**6 (*68*)**	**1 (*100*)**	**8 (93)**

[2] Task duration is zero because the first interaction already started the task execution for the robot.

[3] Due to data corruption Task 1.1 and Task 1.2 duration is not available. Mean value is calculated with three measurements in all cases.

3.2.3 Discussion

This testing aimed to evaluate the new concept of service oriented flexible user interface. No particular selection was in place for the participants and the size of dataset is not wide enough for statistical analysis and true usability evaluation [21]. However, trends and impressions can be synthesized on the difference in user performance. Inspection of the video recordings shows the participants did not understand fully that Task 1 and Task 3 as well as Task 2 and Task 4 were the same except the fact that they have to use different user interface during execution. Although in the series of actions they did not follow the step-by-step instruction of the user manual, participants returned to it over and over again to get acknowledgement on their progression. An excessive number of interactions is present against the number of interactions required by the user manual (Table 6).

Table 6
Comparison of required and performed interactions

	Required	Performed	Difference	Difference [%]
Task 1.1	7	7	0	0%
Task 1.2	9	15	+6	+67%
Task 2	25	42	+17	+68%
Task 3.1	3	6	+3	+100%
Task 3.2	1	1	0	0%
Task 4	6	8	+2	+33%
Total	**51**	**79**	**+28**	**+55%**

At the beginning the state of the user interface and the controller was the same in every test but at the start of Task 1.2 this situation changed due to the different preferences of the users on how to stop a robot program. The significant difference for Task 2 is caused by the fact that there are several ways of inputting a variable in the traditional GUI but the shortest (based upon the robot manufacturer's user manual) was not used by either of the participants.

The excess number of interactions for Task 3 are the actions to dismiss messages on the flexible user interface caused by previous action of the users. The increased number of touches in Task 4 are due to a usability issue: the display item for selecting the amount of parts to be delivered was to small thus the selection could not be made without repeated inputs. Verdict of this investigation is that users tends to use less efficient ways to set up the robot controller which may induce errors and execution time increases due to the need of recuperating from errors.

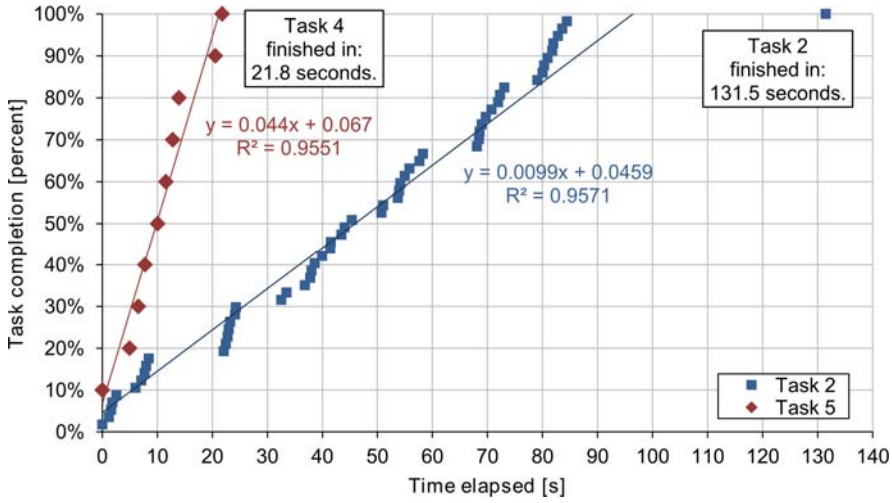


Figure 8
Interaction pattern with traditional (Task 2) and flexible, intuitive system (Task 4)

The intuitiveness of the new approach can be proven with the examination of interaction patterns. Figure 8 shows the interactions of User 1 in details. The user input for the traditional system comes in bursts. The slope of each burst is close to the final slope of the FGUI interaction (Task 4 in Figure 8) but the time between these inputs decreases the overall speed of setting. Recordings show that this time in the case of Task 2 is generally spent on figuring out the next step to execute either by looking in the manual or searching for clues on the display. Finally the press of the start button for Task 2 is delayed because the user double-checked the inputted values.

In contrast the inputs for Task 4 are distributed in time uniformly and the delay between the interactions is significantly lower. The user did not have to refer to the user manual because the user interface itself gave enough information for setup. This means that this composition of user interface which is made specifically for this service of the robot cell offers a more intuitive and easy-to-use interface than the traditional one and that CogInfoCom messages were transmitted successfully.

The overall picture shows that FGUI performs significantly better than TGUI (See Figure 9). For comparison the execution time of Task 3 against Task 1 was shortened by 34 seconds while the duration of Task 4 against Task 2 was shorter by 116 seconds. Regarding the necessary interactions with system FGUI reduced it to 23,4% giving around a quarter of possibilities for errors.

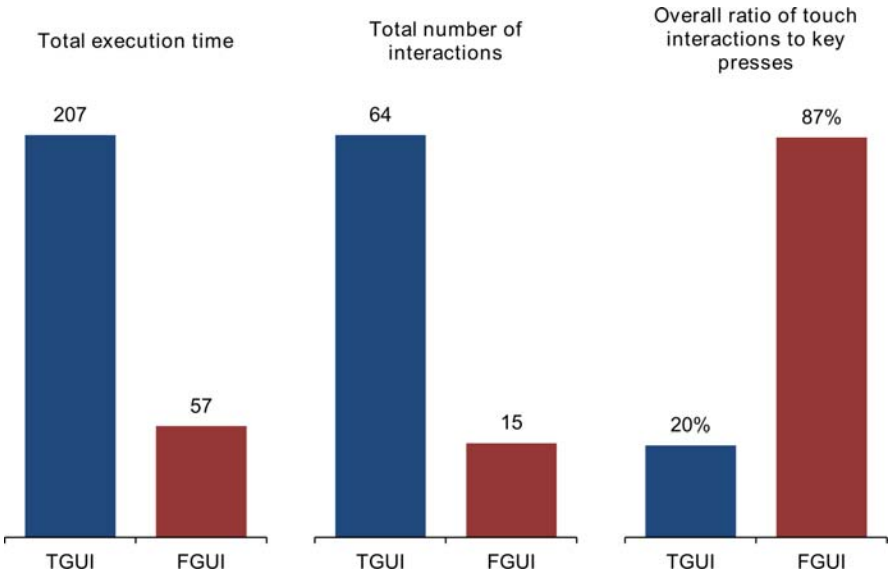


Figure 9
Final results showing improved performance of FGUI: reduced task execution time, decreased number of interaction and increased ratio of touch screen usage to key presses

The Flexible Graphical User Interface also increased the use of touch screen significantly (e.g. from 15% to 93% for Task 2 and Task 4). Participants reported that the use of images and the composition of the UI helped them to connect easier the parameters given by the instructions and the necessary actions to input these values into the controller. This is the result of deliberate design of the abstract level connected to the robot cell's service; the design is based upon the principle of CogInfoCom messages to ensure human centred operation.

Concluding the discussion it is stated that this Flexible Graphical User Interface is evaluated as expected; users were able to operate the robot cell faster, more intuitively and with greater self-confidence. Due to the low number of participants further verification is necessary; organisation of a new test for deeper usability and intuitiveness investigation (including non-expert users) is underway at the time of writing this paper and results will be published in later papers.

Conclusions

The Flexible Graphical User Interface implementation based on Service Oriented Robot Operation was presented. The application of CogInfoCom principles is described and a new property for the CogInfoCom icon notion was introduced. This new property is the role of icon in message transfer and for human robot interaction identification, interaction and feedback roles were identified.

The new approach represented by Service Oriented Robot Operation in human robot interaction for industrial applications was examined by user evaluation. Results show the decrease in task setup and completion time and the reduced number of interactions proves a more intuitive way of communication between man and machine.

Acknowledgement

This work was supported by BANOROB project, the project is funded by the Royal Norwegian Ministry of Foreign Affairs. The authors also would like to thank The Norwegian Research Council for supporting this research work through The Industrial PhD program.

References

[1] J. Scholtz: Theory and evaluation of human robot interactions, System Sciences, 36th Annual Hawaii International Conference on, pp. 10, 2003.

[2] B. Vaughan, J. G. Han, E. Gilmartin, N. Campbell: Designing and Implementing a Platform for Collecting Multi-Modal Data of Human-Robot Interaction, *Acta Polytechnica Hungarica*, Vol. 9, No. 1, pp. 7-17, 2012.

[3] S. Stadler, A. Weiss, N. Mirnig, M. Tscheligi: Anthropomorphism in the factory - a paradigm change?, Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on, pp. 231-232, 2013.

[4] B. Solvang, G. Sziebig: On Industrial Robots and Cognitive Infocommunication, Cognitive Infocommunications (CogInfoCom 2012), 3rd IEEE International Conference on, Paper 71, pp. 459-464, 2012.

[5] World Robotics - Industrial Robots 2011, Executive Summary. Online, Available: http://www.roboned.nl/sites/roboned.nl/files/2011_Executive_Summary.pdf, 2011.

[6] G. A. Boy (ed.): Human-centered automation, in *The handbook of human-machine interaction: a human-centered design approach*. Ashgate Publishing, pp. 3-8, 2011.

[7] Human-centred design for interactive systems, Ergonomics of human-system interaction -- Part 210, ISO 9241-210:2010, International Organization for Standardization, Stand., 2010.

[8] Z. Z. Bien, H.-E. Lee: Effective learning system techniques for human–robot interaction in service environment, Knowledge-Based Systems, Vol. 20, No. 5, pp. 439-456, June 2007.

[9] A. V. Libin, E. V. Libin: Person-robot interactions from the robotpsychologists' point of view: the robotic psychology and robotherapy approach, Proceedings of the IEEE, Vol. 92, No. 11, pp. 1789-1803, Nov. 2004.

[10] P. Korondi, B. Solvang, P. Baranyi: Cognitive robotics and telemanipulation, 15th International Conference on Electrical Drives and Power Electronics, Dubrovnik, Croatia, pp. 1-8, 2009.

[11] FANUC iPendantTM brochure, FANUC Robotics America Corporation (http://www.fanucrobotics.com), 2013.

[12] Yaskawa Motoman Robotics NX100 brochure, Yaskawa America, Inc. (http://www.yaskawa.com), 2011.

[13] FlexPendant - Die neue Version des Programmierhandgertes bietet noch mehr Bedienkomfort, (in German), ABB Asea Brown Boveri Ltd. (www.abb.com), 2009.

[14] KUKA smartPad website, KUKA Roboter GmbH, (www.kuka-robotics.com), 2013.

[15] Smart robot programming, the easy way of programming, Reis GmbH & Co. KG Maschinenfabrik. (www.reisrobotics.de), 2013.

[16] P. Baranyi, B. Solvang, H. Hashimoto, P. Korondi: 3D Internet for Cognitive Infocommunication, Hungarian Researchers on Computational Intelligence and Informatics, 10th International Symposium of 2009 on, pp. 229-243, 2009.

[17] P. Baranyi, A. Csapo: Definition and Synergies of Cognitive Infocommunications, *Acta Polytechnica Hungarica*, Vol. 9, No. 1, pp. 67-83, 2012.

[18] A. Csapo, P. Baranyi: A Unified Terminology for the Structure and Semantics of CogInfoCom Channels, *Acta Polytechnica Hungarica*, Vol. 9, No. 1, pp. 85-105, 2012.

[19] G. Sallai: The Cradle of the Cognitive Infocommunications, *Acta Polytechnica Hungarica*, Vol. 9, No. 1, pp. 171-181, 2012.

[20] B. Daniel, P. Korondi, T. Thomessen: New Approach for Industrial Robot Controller User Interface, Proceedings of the IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society, pp. 7823-7828, 2013.

[21] Wonil Hwang and Gavriel Salvendy: Number of people required for usability evaluation: the 10\±2 rule, *Communication of the ACM*, Vol. 53, No. 5, pp. 130-133, May 2010.

EA Techniques for Optimal Power Flow. Parameter Tuning by Mathematical Test Functions

Florin Solomonesc, Constantin Barbulescu, Stefan Kilyeni, Oana Pop, Marius Cornoiu, Adrian Olariu

"Politehnica" University of Timisoara, 2 Bd. V. Parvan, Timisoara, Romania, claudiu.solomonesc@upt.ro, constantin.barbulescu@upt.ro, stefan.kilyeni@upt.ro, oana.pop@upt.ro, marius.cornoiu@upt.ro, adrian.olariu@upt.ro

Abstract: The goal of this paper is to establish the best values for evolutionary algorithm main parameters. To do this, a real coded algorithm is tested on Rosenbrock, Schwefel and Rastrigin functions. Different genetic operators' implementation ways are described. Several numbers of variables have been analyzed. An original software tool has been developed in a Matlab environment. The most complex three mathematical test functions have been implemented within the software tool; these functions are considered as ideal for studying the behaviour of the algorithm. The settings established are crucial in the optimal power flow computing and the transmission network expansion planning, by means of EA techniques.

Keywords: evolutionary algorithm; test function; crossover rate; mutation rate; random mutation; variable step mutation

1 Introduction

A great amount of attention is paid to Evolutionary Algorithms (EAs) for engineering problem-solving. These algorithms are defined in [1] as population-based meta-heuristic optimization algorithms, having genetics-inspired mechanisms to iteratively refine a set of solution candidates. While the solutions obtained are evaluated according to traditional methods, EAs eliminate complex mathematical computations. Evolutionary Algorithms include the following categories: genetic algorithms (GAs), evolutionary programming (EP) and evolutionary strategies (ESs). The GAs usually use binary digit arrays, whereas the ESs, decimal variable arrays. However, the GAs variables are coded as decimal numbers in several papers.

The main concepts that have been adapted from genetics can be described as follows:

Concept	Meaning in genetics	Mathematical optimization
Phenotype	Set of visible features of an individual; these features are developed on hereditary basis under environmental constraints;	A possible solution;

Concept	Meaning in genetics	Mathematical optimization
Genotype	Set of hereditary proprieties of an organism. Also referred to as a *chromosome*;	A set of variables, put in a form that can be easy processed by the optimization algorithm. usually represented as an array;
Phenome	A group of all phenotypes that define an organ or an organism;	Solution space;
Genome	The group of all different chromosomes;	Search space;
Gene	The basic unit of a chromosome, which carries the hereditary properties;	A decimal or binary variable;
Locus	The position of a gene in a chromosome;	The variable position within an array;
Alela	One of the forms a gene can take;	The variable value;
Population	A set of chromosomes;	A small number of variable sets generated or selected from the search space;
Generation	A population with members of same age.	The population on which the genetic operators are applied during iteration.

Special interest in this kind of optimization methods has been shown in the field of electrical power engineering. A binary coded genetic algorithm for optimal power flow (OPF) is described in [2], based on the FACTS device control (Flexible Alternative Current Transmission Systems). Good results have been obtained for the IEEE14 test power system. [3] describes a binary GA that uses both continuous (real power and bus voltage) and discrete variables (transformer tap ratio) to solve the OPF. The algorithm is tested on IEEE30 and IEEE_3Area_96 systems. [4] proposes a new GA initialization procedure used to determine the OPF. This procedure relies on a voltage grading technique. The OPF problem is also discussed in [5], taking in to account multiple contingencies. A cost efficient OPF method is presented in [6], where the EA uses an adaptive mutation rate. The results refer to the IEEE30 test system, as well as to two real power systems. In [7], the Matlab-integrated genetic algorithm toolbox is used to solve the OPF. The results for the IEEE30 test system are compared to the ones arrived at by means of a particle swarm optimization (PSO) algorithm.

Kazemi et al. [8] put forward a method based on a decimal evolutionary algorithm, to solve the transmission network expansion planning (TNEP). The algorithm has been tested on a modified 6-bus Garver system. In [9], a GA is employed to determine the optimal expansion plan taking into consideration the power loss minimization and the total investment cost for the Azerbaijani power grid. In [10], a method for static TNEP, under deregulation, is presented. The method is based on a GA and a fuzzy decision-making procedure. An improved version of this method, which can deal with dynamic TNEP, is provided in [11]. For the short-term TNEP, a GA combined with a hill-climbing technique is discussed in [12]. Hill-climbing is used for mutation and leads to faster convergence. In [13] and [14], the Pareto dominance is applied to compare the solutions of a multi-objective GA for TNEP.

In [15], the transmission expansion is solved in relation to load uncertainty, with the help of scenario-based GA.

The generation expansion planning (GEP) is also approached using EAs. In [16], the expansion of thermal power plants is discussed taking into account the emissions and their long-term effects. The algorithm is a combination of GA and Bender's decomposition method. in [17] and [18], Murugan et al. apply an NSGA-II (non-dominated sorting genetic algorithm) to solve the GEP for the IEEE30 system. The distribution network expansion is approached by Wang in [19], [20].

The goal of this paper is twofold: to analyze the behaviour of EAs and to determine the optimal values of the main parameters (e.g. population size, mutation and crossover ratios, etc.). The results will be employed in OPF computing and in TNEP, by means of EA techniques. The real coded algorithm is tested on three multiple variable functions: Schwefel, Rastrigin and Rosenbrock. Section 2 describes the implementation of the algorithm. The results are then discussed in section 3.

2 Evolutionary Algorithm (EA) for Test Functions

The flow chart of the proposed EA is presented in Fig. 1. Basically, the algorithm starts from a randomly generated initial population, which becomes the current population for the first iteration (first generation). The genetic operators are applied within each iteration for the current population. The computation process finishes when at least one of the stopping criteria has been fulfilled. In what follows, every step presented in Fig. 1 is described in detail.

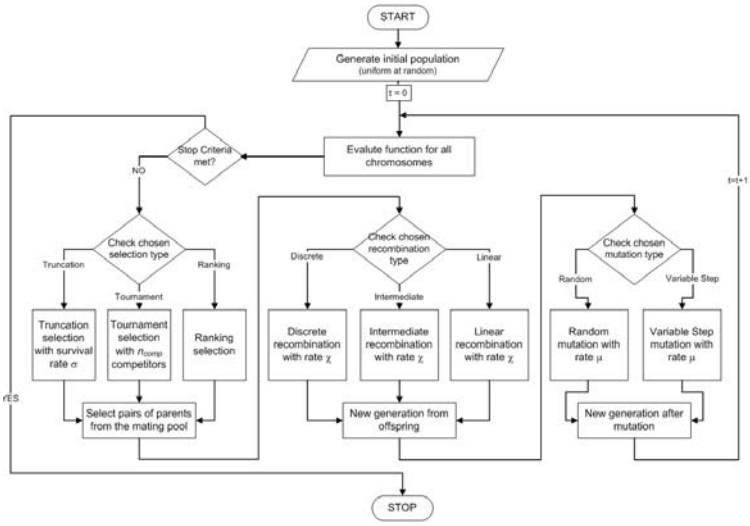


Figure 1
Evolutionary algorithm flow chart

2.1 Variable encoding

The variables are stored within a structure called *chromosome*, which is represented as an array. Each variable, whether discrete or continuous, occupies a number of elements from the array (these elements are called *genes*). If binary representation is used, then the number of genes for each variable is set according to the function domain and the desired precision. In this paper, real representation is used, so every gene represents a variable. The chromosome can be represented as the set $\boldsymbol{x} = \{x_1, x_2, ..., x_d\}$, where d stands for the number of variables. Fig. 2 shows an example of such a chromosome, used to optimize the Schwefel function with six variables. According to [21], this function can take any number of variables and is defined within the [-500; 500] interval.

-61,2556	251,2671	-148,34	-337,818	-393,347	353,0311

Figure 2
Chromosome structure example

2.2 Population size and initial population

Several chromosomes (individuals) form the population. A t moment population is called a *generation*, where t stands for the generation counter. The population size n_c, which is considered as fixed during one run of the algorithm, is one of the parameters that are studied within this paper. The initial population $\boldsymbol{x}^0$ is determined using Eq. (1).

$$\begin{aligned} &\boldsymbol{x}^0_{i,j} = (x_{max} - x_{min}) \cdot a_{i,j} + x_{min} \qquad i = 1, 2, ..., n_c \\ &a_{i,j} \in [0;1], \text{ uniform at } random \qquad j = 1, 2, ..., d \end{aligned} \tag{1}$$

As it can be seen from Eq. (1), the initial population has a uniform random distribution between the domain limits [x_{min}; x_{max}] and is generated without other constraints. It may be necessary to consider several constraints when generating the initial population, so that the algorithm starts from a set of feasible solutions.

2.3 Evaluation and selection

The function value is evaluated for each chromosome within the population. Thus, for each $\boldsymbol{x}_i$ array, there will be a $f_i = f(\boldsymbol{x}_i)$ value. In engineering, most of the optimization problems are minimization problems, the smallest f_i value being the best.

Selection ensures that the best fitted individuals have a chance to produce offsprings for the next generation, but it also offers a small chance to the others to be selected as well. The individuals are first selected according to the function values and then copied in a buffer matrix called *the mating pool* (***MP***). Three selection types have been implemented: tournament, truncation and ranking.

Truncation selection is very easy to apply. The corresponding parameter is the *survival rate* σ and it represents the population percentage that will be selected for

reproduction. This parameter is conventionally inputted as a number between 0 and 1 (e.g. 0.5 represents half of the population). If the survival rate is low (e.g. 0.25), there is a risk of losing diversity and if it is high (e.g. 0.75), convergence may be slow.

The stages of this selection are the following:

- the population is *sorted ascending* according to the function values;
- the best $n_{sel} = \sigma \cdot n_c$ chromosomes are copied in the mating pool;
- copies of these chromosomes are made until the mating pool is full (the number of chromosomes in the mating pool equals n_c).

Given that the truncation selection is an iterative process, the following relations may be established for iteration k:

$$\begin{aligned} & f(\boldsymbol{x}_{sel}^k) = \min\{f(\boldsymbol{x}_i^k)\} \qquad i = 1,2,...,n_c + 1 - k, \\ & \boldsymbol{MP}^k = \boldsymbol{MP}^{k-1} \cup \boldsymbol{x}_{sel}^k \qquad k = 1,2,...,n_{sel} \\ & \boldsymbol{P}^{k+1} = \boldsymbol{P}^k \setminus \boldsymbol{x}_{sel}^k \end{aligned} \tag{2}$$

Tournament selection finds the fittest individual in a small group randomly selected from the current population. The parameter n_{comp} designates the number of competitors (the size of the group of chromosomes to be compared).

This is an iterative process, which comprises the following stages:

- a group of n_{comp} chromosomes is randomly chosen form the current population ($\boldsymbol{x}_{comp_i}$, i = 1,2,...,n_{comp});
- the function values for each chromosome are compared to find their minimum;
- the chromosome corresponding to the minimum function value is copied in the mating pool;
- the process stops if the size of the mating pool equals n_c.

The following relations may be written for iteration k:

$$\begin{aligned} & a_i^k \in \{1,2,...,n_c\}, \text{ at random} \\ & \boldsymbol{x}_{comp_i}^k = \boldsymbol{x}_{a_i}^k \qquad i = 1,2,...,n_{comp} \\ & f(\boldsymbol{x}_{sel}^k) = \min\{f(\boldsymbol{x}_{comp_i}^k)\} \qquad k = 1,2,...,n_c \\ & \boldsymbol{MP}^k = \boldsymbol{MP}^{k-1} \cup \boldsymbol{x}_{sel}^k \\ & \boldsymbol{P}^{k+1} = \boldsymbol{P}^k \end{aligned} \tag{3}$$

Note that a chromosome can be selected several times. If parameter n_{comp} has a high value, it may lose diversity, because a good chromosome is chosen too many times. However, if it is too low, suitable chromosomes may be omitted in the selection process.

Ranking selection is derived from roulette wheel selection. Unlike the latter, it can manage negative function values, by working with *ranks* instead. For a better understanding, a detailed description of each stage of this process is presented below:

- the population is *sorted descending* according to the function values;
- each chromosome in the sorted population is ranked from *1* to n_c - *1* being the worst and being n_c best:

$$rank_i = rank(\boldsymbol{x}_i) = i \qquad i = 1, 2, ..., n_c \tag{4}$$

- the selection probability of a chromosome is computed:

$$p_i = p(\boldsymbol{x}_i) = rank_i / \sum_i rank_i \qquad i = 1, 2, ..., n_c \tag{5}$$

- the cumulative sum of these probabilities is expressed as:

$$Cs_i = Cs(\boldsymbol{x}_i) = \sum_{j=1}^{i} p_j \qquad i = 1, 2, ..., n_c \tag{6}$$

- a number is randomly generated within the interval [*0*; *1*); the first chromosome that has the corresponding cumulative sum of probabilities greater than the random number is copied in the mating pool;
- the process stops if the size of the mating pool equals n_c.

As in the previous case, a chromosome can be selected several times. This is an iterative process. The following relations may be written for iteration *k*:

$$\begin{aligned} &a \in [0;1], \text{ uniform and random} \\ &Cs(\boldsymbol{x}_{sel}^k) = \min\{Cs(\boldsymbol{x}_i^k) \mid Cs(\boldsymbol{x}_i^k) > a\} \qquad i = 1, 2, ..., n_c \\ &\boldsymbol{MP}^k = \boldsymbol{MP}^{k-1} \cup \boldsymbol{x}_{sel}^k \qquad k = 1, 2, ..., n_c \\ &\boldsymbol{P}^{k+1} = \boldsymbol{P}^k \end{aligned} \tag{7}$$

2.4 Mating and Crossover

The next step in the algorithm is to create pairs of parents (mating) from the chromosomes copied in the mating pool. The pairs are chosen by randomly selecting two different positions in the mating pool until $n_c / 2$ pairs are created. Various situations may arise: a chromosome is never chosen to mate; a chromosome is chosen to mate several times; if a chromosome was selected several times, it may mate with another copy of itself. Not all the pairs of parents will undergo crossing-over; some will be copied in the new generation as they are. The number of pairs that will undergo crossing-over is given by the *crossover rate* χ, which is conventionally considered as a number between 0 and 1. In this paper, the crossover rate ranges from 0.5 to 0.9.

Crossing-over generates two new sets of variables from two sets of parent variables, based on a mathematical relation. Three crossover types are discussed in this

paper: discrete (uniform), intermediate and linear. The number of pairs that undergo crossing-over is $n_{rp} = \chi \cdot n_c / 2$, irrespective of the type used.

All the three crossover types are based on relations 8 [22]

$$\begin{aligned} \boldsymbol{x}^{o1} &= \left\{ r \cdot x_i^M + (1-r) \cdot x_i^F \right\} \\ \boldsymbol{x}^{o2} &= \left\{ r \cdot x_i^F + (1-r) \cdot x_i^M \right\} \end{aligned} \qquad i = 1, 2, ..., d \tag{8}$$

where $\boldsymbol{x}^M, \boldsymbol{x}^F$ – parents and $\boldsymbol{x}^{o1}, \boldsymbol{x}^{o2}$ – offsprings.

Variable *r* is randomly generated. The way it is represented as well as the sets and intervals of its values defines the crossover type. For *discrete crossover*, *r* is a *2 x d* matrix containing zeros and ones, as illustrated by the following relation:

$$\boldsymbol{r} = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,d} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,d} \end{bmatrix}, \qquad r \in \{0,1\}, \text{ at random} \tag{9}$$

Relation 8 thus becomes:

$$\begin{aligned} \boldsymbol{x}^{o1} &= \left\{ r_{1,i} \cdot x_i^M + (1-r_{1,i}) \cdot x_i^F \right\} \\ \boldsymbol{x}^{o2} &= \left\{ r_{2,i} \cdot x_i^F + (1-r_{2,i}) \cdot x_i^M \right\} \end{aligned} \qquad i = 1, 2, ..., d \tag{10}$$

Intermediate crossover produces the most changes in the population. In this case, *r* is a 2x*d* matrix containing random values between 0 and 1, as described by relation (11). Relation 10 can be used to calculate the values of the offspring variables.

$$\boldsymbol{r} = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,d} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,d} \end{bmatrix}, \qquad r \in [0;1], \text{ uniform and random} \tag{11}$$

In the case of *linear crossover*,e, $\boldsymbol{r}$ is a 2x1 matrix containing random values between 0 and 1.

$$\boldsymbol{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}, \qquad r \in [0;1], \text{ uniform and random} \tag{12}$$

Relation 8 thus becomes:

$$\begin{aligned} \boldsymbol{x}^{o1} &= \left\{ r_1 \cdot x_i^M + (1-r_1) \cdot x_i^F \right\} \\ \boldsymbol{x}^{o2} &= \left\{ r_2 \cdot x_i^F + (1-r_2) \cdot x_i^M \right\} \end{aligned} \qquad i = 1, 2, ..., d \tag{13}$$

2.5 Mutation

Mutation in EA mimics the natural process by changing the values of a small number of randomly selected genes. Mutation is subsequently applied after crossover and the authors refer to both as *reproduction*.

Mutation is performed according to the mutation rate μ, which represents the percentage of genes that are subjected to mutation. The mutation rate conventionally ranges between 0 and 1. Therefore, the number of mutated genes is $n_{mg} = \mu \cdot d \cdot n_c$.

In this paper, the mutation rate ranges from 0.05 to 0.5.

For real coded EA, mutation can be performed by replacing a randomly chosen gene from a randomly chosen chromosome with a randomly generated value in the function domain. This method is addressed to as *random mutation*. For any chromosome $\boldsymbol{x}_i$, the position of a gene to be mutated is considered ℓ_{mut}. The new value of the gene is calculated as follows:

$$\begin{aligned} &\boldsymbol{x}_{i,\ell mut} = (x_{max} - x_{min}) \cdot a + x_{min}, \quad i \in \{1,2,...,n_c\} \\ &a \in [0;1], \text{ uniform ant random} \end{aligned} \tag{14}$$

The authors have also tested another kind of mutation – variable step mutation. For any chromosome $\boldsymbol{x}_i$ and position ℓ_{mut} of a gene to be mutated, the new value of the gene is established according to the relation:

$$\begin{aligned} &x_{i,\ell mut} = x_{i,\ell mut} \pm (x_{max} - x_{min}) \cdot a \cdot step \\ &a \in [0;1], \text{ uniform at random} \end{aligned} \tag{15}$$

The value of a gene is altered by a quantity proportional to the domain. The variable step represents that proportion and it is divided by 10, after a number of iterations, to refine the current solution. The initial step size is 0.1. This particular mutation type may cause a variable to exceed the domain. If this case, the variable is assigned the value of the limit it has exceeded.

2.6 Elitism

Elitism implies that a small number of the most suitable solutions are copied unaltered to the next generation. The approach put forth in this paper copies only one solution (the best), as described by Eq. 16.

$$\begin{aligned} &f(\boldsymbol{x}_{elit}^t) = \min\{f(\boldsymbol{x}_i^t)\}, \quad i = 1,2,...,n_c \\ &\boldsymbol{x}_1^{t+1} = \boldsymbol{x}_{elit}^t \end{aligned} \tag{16}$$

2.7 Stopping criteria

Once a solution has been found, one has determine the extent to which this solution provides a fair answer to the problem. The best solution is usually a compromise between quality and computational effort. Consequently, EA performance has been analyzed under different conditions, considering the following two stopping criteria:

- the quality of the solution no longer improves after a certain number of iterations;
- the maximum number of generations is reached (backup criterion).

3 Results and Discussion

Based on the above algorithm, a software tool has been developed in MatLab 2012. This software tool allows the user to change the values of basic parameters, such as population size, crossover rate and mutation rate. Moreover, the user can change between the different selection, reproduction and mutation types, and can set specific parameters.

The EA settings are the following:

- the number of variables for each function is 10;
- truncation selection is used, with $\sigma = 0.5$ survival rate (as a result, the best individuals are selected in a reduced computing time);
- the intermediate crossover method is used, which ensures a high population diversity is assured;
- both mutation methods are tested. In the case of variable step mutation, the step decreases every 200 iterations;
- the maximum iteration number is 5000;
- for each situation, the algorithm runs 20 times.

The influence of the parameters is analyzed for the following values:

- population size: $n_c = \{20; 40; 60; 80; 100\}$;
- crossover rate: $\chi = \{0.5; 0.6; 0.7; 0.8; 0.9\}$;
- mutation rate: $\mu = \{0.01; 0.05; 0.1; 0.25; 0.5\}$.

Once the optimal parameters have been established for each function, the provided values are presented for different configurations of genetic operators. The cases under discussion are summarized in Table 1. When not subjected to analysis, the size of the population counts 20 individuals, and the crossover probability is 0.8. Mutation probability is set after running a test.

Table 1
Cases under discussion for EA testing

	EA1	EA2	EA3	EA4	EA5
Selection:	Truncation	Tournament	Ranking	Truncation	Truncation
• Survival rate:	$\sigma = 0.5$	-	-	$\sigma = 0.5$	$\sigma = 0.5$
• Number of competitors:	-	$n_{comp} = n_c/4$	-	-	-
Crossover	Intermediate	Intermediate	Intermediate	Discrete	Linear

Three mathematical test functions have been implemented within the software tool: Schwefel, Rastrigin and Rosenbrock. The definitions of these functions, the intervals of their variables and their three-dimensional representation are presented below.

3.1 Rastrigin function

The Rastrigin function is a nonlinear function (relation (17)). It is based on De Jong function, additionally including the cosine term, which periodically generates local minimum values with regulated distribution [21]. Due to the great number of local minimum values, the optimization EAs have difficulties in finding the global minimum [23]. The corresponding 3D plot is presented in Fig. 3. The global optimum value is recorded for $x = 0$ ($f(x) = 0$).

$$f = 10 \cdot n + \sum_{i=1}^{n} (x_i^2 - 10 \cdot \cos(2\pi \cdot x_i)), \quad i = \overline{1,n}\ ; \quad -5.12 \le x_i \le 5.12 \tag{17}$$

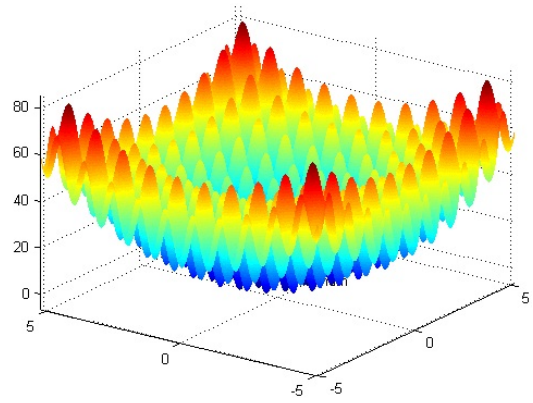


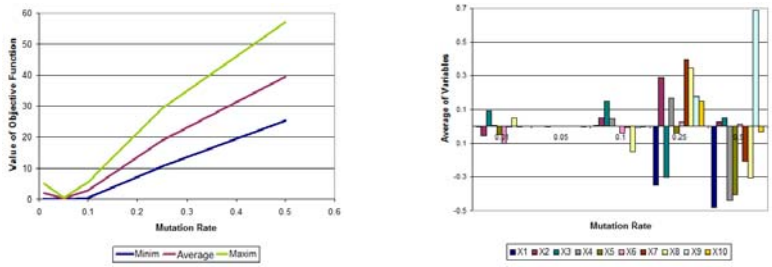
Figure 3
3D Rastrigin's plot

The evolution of minimum, average and maximum function values are presented in Fig. 4.a. The average variable values are presented in Fig. 4.b. Random mutation for a 0.05 rate constantly leads to very good results. Variable step mutation produces inadequate results (Fig. 5), even for reduced mutation rates.

In what follows, a 0.05 mutation rate will be used.

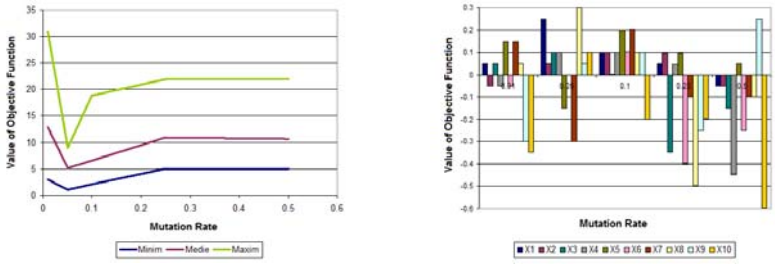
Fig. 6 describes the influence of the crossover rate. For a 0.7 or a 0.8 rate, the solutions are concentrated in the neighbourhood of the global minimum. Fig. 7 illustrates the population influence. It should be emphasized that a large population size has a negative effect on the quality of the solution.

The numerical results corresponding to the Rastrigin function, algorithms EA1-EA5 (Table 1), are summarized in Table 2. The settings used in the analyses are: population size $n_c = 20$; crossover rate $\chi = 0.8$; mutation type – random; mutation rate $\mu = 0.05$.



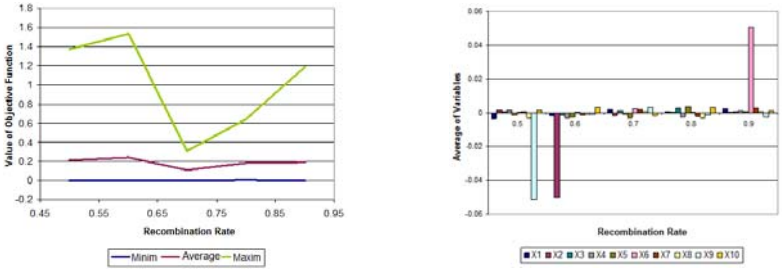
a) b)

Figure 4
Mutation rate influence – random mutation; a) function value; b) average variables' values



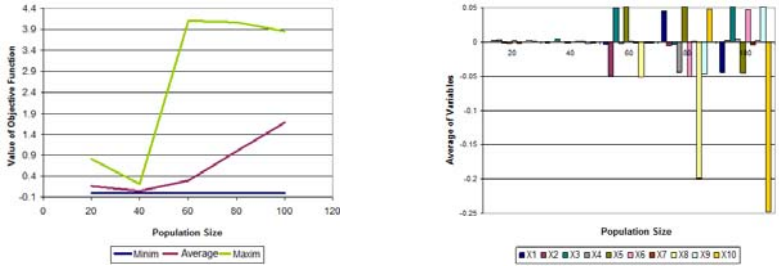
a) b)

Figure 5
Mutation rate influence – variable step mutation; a) function value; b) average variables' values



a) b)

Figure 6
Crossover rate influence a) function value; b) average variables' values



a) b)

Figure 7
Population size influence a) function value; b) average variables' values

Table 2
Rastrigin function numerical results' synthesis

	EA1	EA2	EA3	EA4	EA5
$f(\mathbf{x}_i)$ average value	0.1509	0.1714	0.5227	0.2596	0.2847
$f(\mathbf{x}_i)$ minimum value	0.0094	0.0012	0.0221	0.0147	0.0039
x_1 average value	-0.0002	0.0000	0.0035	0.0000	0.0035
x_2 average value	0.0000	0.0029	-0.0001	0.0029	-0.0001
x_3 average value	-0.0009	0.0004	-0.0022	0.0004	-0.0022
x_4 average value	-0.0006	-0.0016	0.0392	-0.0016	0.0392
x_5 average value	-0.0002	-0.0015	-0.0189	-0.0015	-0.0189
x_6 average value	0.0004	-0.0012	-0.0000	-0.0012	-0.0000
x_7 average value	0.0007	-0.0001	-0.0018	-0.0001	-0.0018
x_8 average value	0.0017	0.0004	-0.0204	0.0004	-0.0204
x_9 average value	-0.0005	-0.0007	0.0189	-0.0007	0.0189
x_{10} average value	0.0011	0.0000	-0.0008	0.0000	-0.0008

These results are graphically presented in Fig. 8.

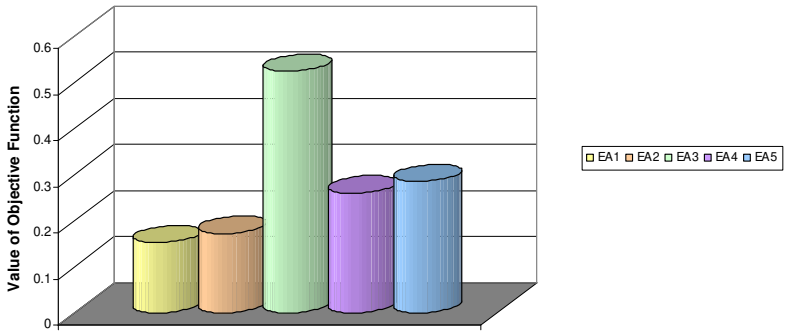


Figure 8
Rastrigin function values in case of each algorithm EA1-EA5

Fig. 8 and Table 2 highlight the following aspects:

- EA1 produces the best results (truncation selection and intermediate crossover);
- the average function values range from 0.1509 to 0.5227;
- the average variable values are range from -0.0204 to 0.0392; both limits are recorded for EA3 (ranking selection and intermediate crossover);
- truncation selection induces high quality behaviour, if associated with powerful crossover methods;
- compared to the EA3, other combinations lead to insignificant variations when the number of variables is increased.

3.2 Schwefel function

The Schwefel function is a nonlinear function (relation (18)). This function in characterized by an increased distance between the global minimum and the subsequent one [21], could lead to a wrong direction of the EA [23]. The corresponding 3D plot is presented in Fig. 9. The global optimum value is recorded for $x = 420.9687$ ($f(x) = 0$).

$$f = 418,9829 \cdot n - \sum_{i=1}^{n} [-x_i \cdot \sin(\sqrt{|x_1|}), \quad i = \overline{1,n}\ ; \quad -500 \le x_i \le 500 \tag{18}$$

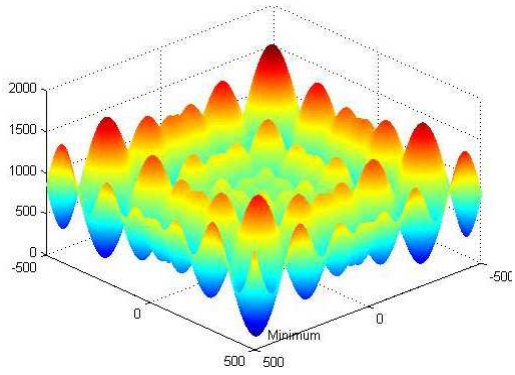
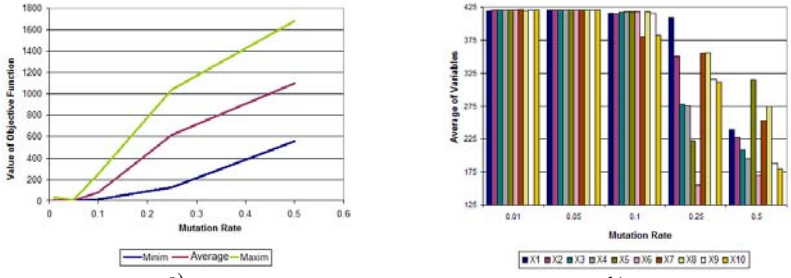


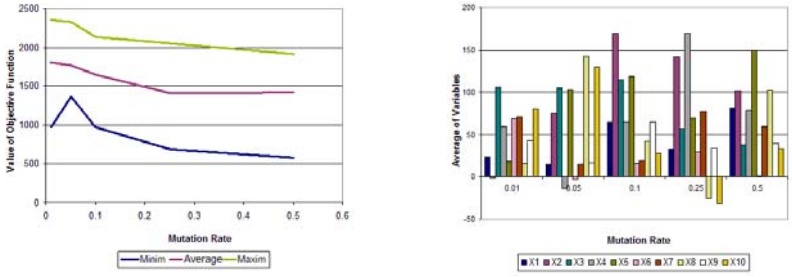
Figure 9
3D Schwefel's plot

The evolution of minimum, average and maximum function values are presented in Fig. 10.a. The average variable values are presented in Fig. 10.b. Random mutation for a 0.05 rate constantly leads to very good results.



a) b)

Figure 10
Mutation rate influence – random mutation; a) function value; b) average variables' values

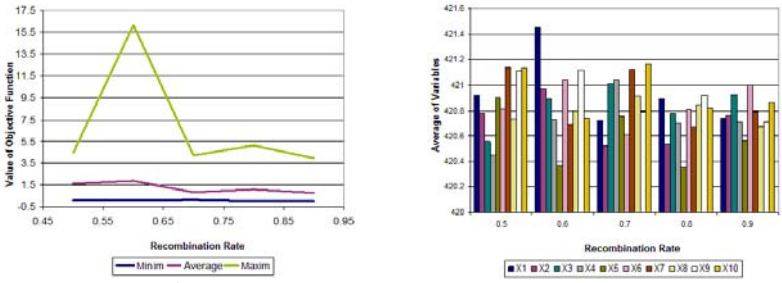


a) b)

Figure 11
Mutation rate influence – variable step mutation; a) function value; b) average variables' values

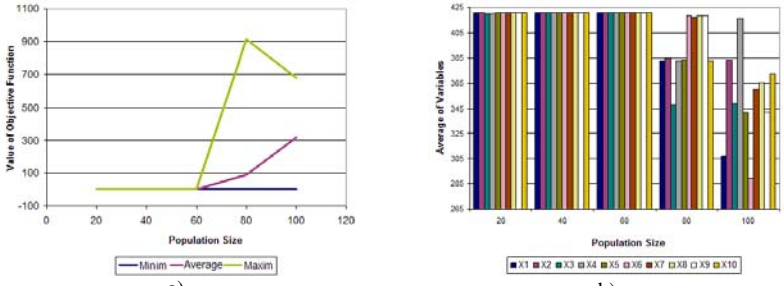
In what follows, a 0.05 mutation rate will be used.

Fig. 12 describes the influence of the crossover rate. The solution quality is improving for high crossover rate values. Fig. 13 illustrates the population influence. It should be emphasized that a large population size has a negative effect on the quality of the solution.



a) b)

Figure 12
Crossover rate influence: a) function value; b) average variables' values



a) b)

Figure 13
Population size influence: a) function value; b) average variables' values

The numerical results corresponding to the Schwefel function, algorithms EA1-EA5 (Table 1) are summarized in Table 3. The settings used in the analyses are: population size $n_c = 20$; crossover rate $\chi = 0.7$; mutation type – random; mutation rate $\mu = 0.05$.

Table 3
Schwefel function numerical results' synthesis

	EA1	EA2	EA3	EA4	EA5
$f(\mathbf{x}_i)$ average value	0.7936	0.6936	22.6335	1.7793	1.6594
$f(\mathbf{x}_i)$ minimum value	0.0293	0.0053	0.8711	0.1681	0.0281
x_1 average value	420.82	420.88	418.85	420.80	420.96
x_2 average value	421.02	420.86	418.46	420.82	420.78
x_3 average value	420.74	420.92	419.34	420.84	420.60
x_4 average value	420.94	420.64	417.94	421.00	420.66
x_5 average value	420.70	421.00	418.96	420.90	420.68
x_6 average value	420.90	420.98	419.18	420.76	420.90
x_7 average value	420.80	420.82	418.74	421.02	420.98
x_8 average value	421.00	421.08	417.68	420.82	420.78
x_9 average value	420.86	420.96	418.78	421.18	420.66
x_{10} average value	420.87	420.87	418.00	421.09	420.67

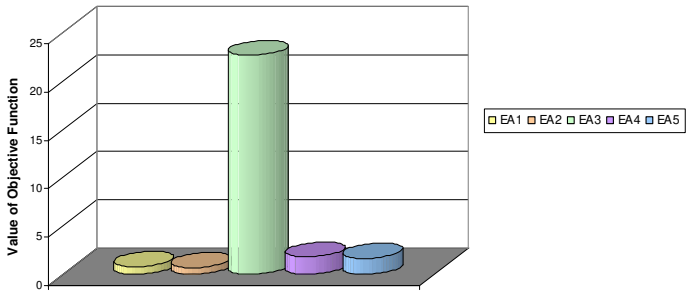


Figure 14
Schwefel function values in case of each algorithm EA1-EA5

Fig. 14 and Table 3 highlight the following aspects:

- EA2 produces the best results (tournament selection and intermediate crossover);
- the average function values range from 0.6936 to 22.6335;
- the average variable values are range from 420.64 to 421.08; both limits are recorded for EA3 (ranking selection and intermediate crossover);
- truncation selection and tournament selection induce high quality behaviour, if associated with powerful crossover methods;
- other combinations lead to insignificant variations when the number of variables is increased.

3.3 Rosenbrock function

Rosenbrock function (relation (19)) represents a classical optimization problem. The global optimum is situated inside a long and narrow valley. The searching algorithms reache the valley easily, but the convergence to the global optimum is difficult. This function is used fortest test the performance of the searching algorithm [21]. The corresponding 3D plot is presented in Fig. 15. The global optimum value is recorded for $x = 1$, $f(x) = 0$.

$$f = \sum_{i=1}^{n-1} [100 \cdot (x_{i+1} - x_i^2)^2 + (1 - x_i)^2] ; \qquad -2{,}048 \le x_i \le 2{,}048 \tag{19}$$

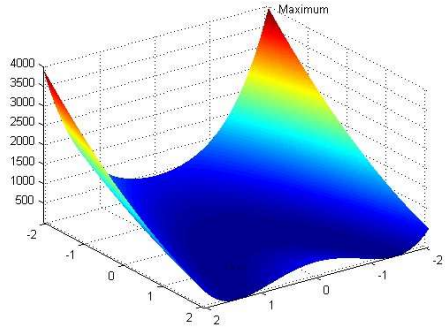
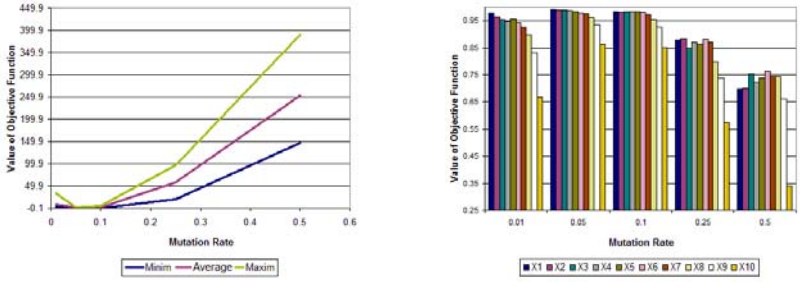


Figure 15
Rosenbrock function 3D plot

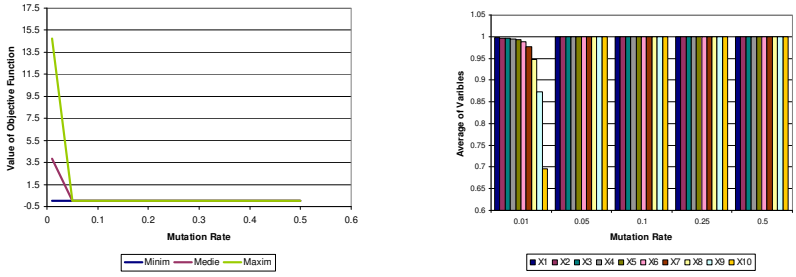
The variable step mutation proves to be the most suitable mutation method for determining the global minimum of this function (Fig. 16).



a) b)

Figure 16
Mutation rate influence – random mutation; a) function value; b) average variables' values

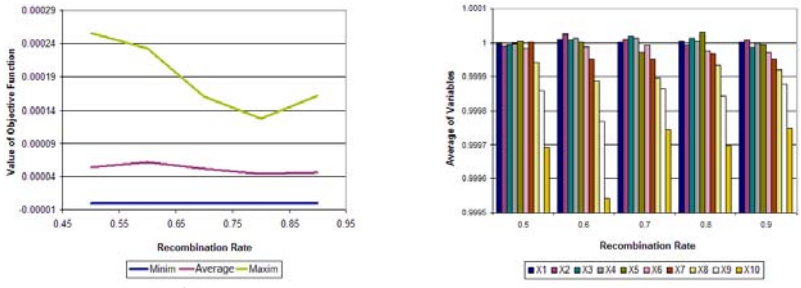
Variable step mutation with mutation rates greater than 0.05 leads to the right solution, as seen in Fig. 17. In what follows, a 0.25 mutation rate will be used.



a) b)

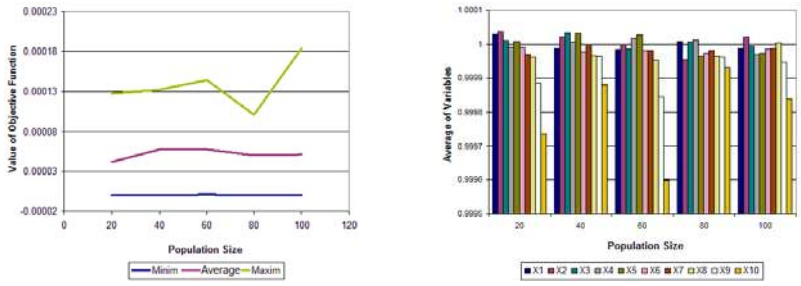
Figure 17
Mutation rate influence – random mutation; a) function value; b) average variables' values

Fig. 18 describes the influence of the crossover rate. Fig. 19 illustrates the population influence. It should be emphasized that their influence is not significant on the quality of the solution.



a) b)

Figure 18
Crossover rate influence: a) function value; b) average variables' values



a) b)

Figure 19
Population size influence: a) function value; b) average variables' values

The numerical results corresponding to the Rosenbrock function, algorithms EA1-EA5 are summarized in Table 4. The settings used in the analyses are: population size $n_c = 20$; crossover rate $\chi = 0.8$; mutation type – variable step; mutation rate $\mu = 0.25$.

Table 4
Rosenbrock numerical results' synthesis

	EA1	EA2	EA3	EA4	EA5
$f(\mathbf{x}_i)$ average value	0.0001	0.0001	0.0000	0.0001	0.0001
$f(\mathbf{x}_i)$ minimum value	0	0	0	0	0
x_1 average value	1.0000	1.0001	1.0000	1.0000	0.9999
x_2 average value	1.0000	1.0000	1.0000	0.9999	1.0000
x_3 average value	1.0000	1.0000	1.0000	0.9999	0.9999
x_4 average value	1.0000	1.0000	0.9999	1.0000	0.9999
x_5 average value	1.0000	1.0000	0.9999	0.9999	0.9999
x_6 average value	0.9999	0.9999	0.9999	0.9999	0.9999
x_7 average value	1.0000	0.9999	0.9999	0.9999	0.9999
x_8 average value	0.9999	0.9999	0.9999	0.9999	0.9999
x_9 average value	0.9999	0.9999	0.9999	0.9998	0.9999
x_{10} average value	0.9999	0.9998	0.9999	0.9998	0.9998

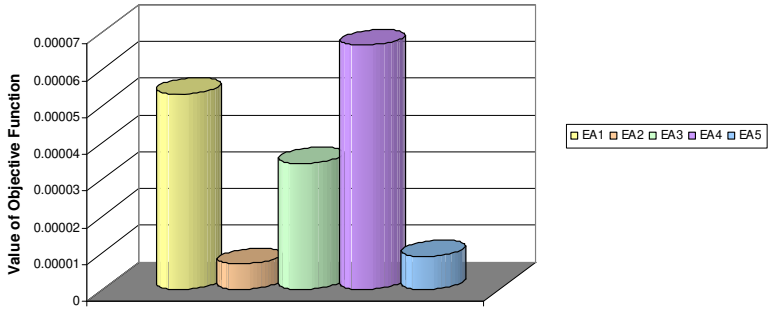


Figure 20
Rosenbrock function values in case of each algorithm EA1-EA5

Fig. 20 and Table 4 highlight the following aspects:

- EA2 produces the best results (tournament selection, intermediate crossover);
- the average function values range from 0.0000 to 0.0001;
- the average variable values are range from 0.9998 to 1.0001;

- EA2 and EA5 algorithms are constantly producing very good results;
- variable step mutation having high mutation rate could be considered the main mechanism.

3.4 A contrastive analysis of the three functions

Table 5 shows the best parameter values for the three functions discussed above.

Table 5
Results and settings for 20, 50 and 100 variables

Function	Variables	Population size	Mutation type	Mutation rate	Crossover rate	Function value	Variables' average value
Rastrigin	20	50	Random	0.01	0.8	0.0029	-0.0005
Rastrigin	50	100	Random	0.01	0.8	0.0170	0.0001
Rastrigin	100	200	Random	0.01	0.8	2.9807	0.0005
Schwefel	20	20	Random	0.01	0.7	0.0074	420.9936
Schwefel	50	100	Random	0.01	0.7	0.0039	420.9975
Schwefel	100	200	Random	0.01	0.7	1.7627	420.9068
Rosenbrock	20	20	Variable step	0.25	0.8	0.0000	0.9999
Rosenbrock	50	50	Variable step	0.25	0.8	0.0003	0.9999
Rosenbrock	100	100	Variable step	0.25	0.8	0.0158	0.9996

The computing process stops if during 200 iterations the result does not improve or the maximum number of iterations is reached (in this case – 5000 iterations). Truncation selection produces good results, characterized by reduced computational effort. A 0.5 survival rate has been used. Intermediate crossover produces the greatest population diversity.

Conclusion

The present research is not only necessary, but also extremely useful step to further developments in EA based OPF computing and TNEP analysis.

In this paper the authors have attempted to discuss the behaviour of the EAs in several scenarios. A suitable software tool has been developed for the EAs benchmark. Different combinations between the genetic operators and their values have been analyzed for several number of variables. The results can be applied to power systems analysis.

The values of the EA parameters are directly influenced by the nature of the optimization problem and by the number of variables.

For a small number of variables, ranking selection leads to very good results.

It is recommended that the population size should not to be too large.

High rate values are recommended for crossover, while reduced rate values for random mutation. If variable step mutation is used, high mutation rate values are can be employed (0.05-0.5). A 0.25 value has been adopted for the analyzed cases. This strategy leads to a decrease in population size. Such an algorithm is very close to the evolutionary computing strategy.

Variable step mutation offers the best results for Rosenbrock function, when used with high probability. Rastrigin and Schwefel functions can be solved using random mutation and a low mutation rate.

If variable step mutation is employed, the influence of the crossover rate and of the population size is less significant; the algorithm is mutation driven.

Satisfactory results have been obtained for a bigger number of variables

Acknowledgement

This work was supported by the strategic grant POSDRU/89/1.5/S/57649, Project ID 57649 (PERFORM-ERA), co-financed by the European Social Fund – Investing in People, within the Sectoral Operational Programme Human Resources Development 2007-2013 and by the strategic grant POSDRU/ 88/1.5/S/50783 (2009) of the Ministry of Labour, Family and Social Protection, Romania, co-financed by the European Social Fund – Investing in people.

References

[1] Weise T., Global optimization algorithms – theory and application, 2nd Edition, http://www.it-weise.de/, 2010;

[2] Chung T.S., Li Y.Z., A Hybrid GA Approach for OPF with Consideration of FACTS Devices, IEEE Power Engineering Review, vol.21, no.2, 2001, pp.47-50;

[3] Bakirtzis A. G., Biskas P. N., Zoumas C. E., Petridis V., Optimal Power Flow by Enhanced Genetic Algorithm, IEEE Power Engineering Review, vol.22, no.2, 2002, pp.60-66;

[4] Todorovski M., Rajicic D., An initialization procedure in solving optimal power flow by genetic algorithm, IEEE Transactions on Power Systems, vol.21, no.2, 2006, pp.480-487;

[5] Chan K.Y., Ling S.H., Chan K.W., Iu H.H.C., Pong G.T.Y., Solving multi-contingency transient stability constrained optimal power flow problems with an improved GA, IEEE Congress on Evolutionary Computation, CEC 2007, pp.2901-2908;

[6] Malik I.M., Srinivasan D., Optimum power flow using flexible genetic algorithm model in practical power systems, 9th International Power and Energy Conference IPEC 2010, pp.1146-1151;

[7] Rahul J., Sharma Y., Birla D., A New Attempt to Optimize Optimal Power Flow Based Transmission Losses Using Genetic Algorithm, 4th International Conference on Computational Intelligence and Communication Networks, CICN 2012, pp.566-570;

[8] Kazemi A., Jalilzadeh S., Mahdavi M., Haddadian H., Genetic algorithm-based investigation of load growth factor effect on the network loss in TNEP, 3rd IEEE Conference on Industrial Electronics and Applications ICIEA 2008, pp.764-769;

[9] Jalilzadeh S., Kazemi A., Mahdavi M., Haddadian H., TNEP considering voltage level, network losses and number of bundle lines using GA, 3rd International Conference Electric Utility Deregulation and Restructuring and Power Technologies DRPT 2009, pp.1580-1585;

[10] Maghouli P., Hosseini S.H., Buygi M.O., Shahidehpour M., A Multi-Objective Framework for Transmission Expansion Planning in Deregulated Environments, IEEE Transactions on Power Systems, vol.24, no.2, 2009, pp.1051-1061;
[11] Maghouli P., Hosseini S.H., Buygi M.O., Shahidehpour M., A Scenario-Based Multi-Objective Model for Multi-Stage Transmission Expansion Planning, IEEE Transactions on Power Systems, vol.26, no.1, 2011, pp.470-478;
[12] Rodriguez J., Falcao D.M., Taranto G.N., Almeida H., Short-Term Transmission Expansion Planning by a Combined Genetic Algorithm and Hill-Climbing Technique, 15th International Conference on Intelligent System Applications to Power Systems, ISAP 2009, pp.1-6;
[13] F. Cadini, E. Zio, C.A. Petrescu, Optimal expansion of an existing electrical power transmission network by multi-objective genetic algorithms, Reliability Engineering & System Safety, Volume 95, Issue 3, 2010, pp. 173-181;
[14] Huang Wei, Feng Li, He Zijun, Cui Junzhao, Zhang Li, Transmission network planning with N-1 security criterion based on improved multi-objective genetic algorithm, 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies DRPT 2011, pp.1250-1254;
[15] Asadzadeh V., Golkar M.A., Moghaddas-Tafreshi S.M., Economics-based transmission expansion planning in restructured power systems using decimal codification genetic algorithm, IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies AEECT 2011, pp.1-8;
[16] Sirikum J., Techanitisawad A., Kachitvichyanukul V., A New Efficient GA-Benders' Decomposition Method: For Power Generation Expansion Planning With Emission Controls, IEEE Transactions on Power Systems, vol.22, no.3, 2007, pp.1092-1100;
[17] Murugan P., Kannan S., Baskar S., Application of NSGA-II Algorithm to Single-Objective Transmission Constrained Generation Expansion Planning, IEEE Transactions on Power Systems, vol.24, no.4, 2009, pp.1790-1797;
[18] Kannan S., Baskar S., McCalley J.D., Murugan P., Application of NSGA-II Algorithm to Generation Expansion Planning, IEEE Transactions on Power Systems, vol.24, no.1, 2009, pp.454-461;
[19] Wang D.T., Ochoa L.F., Harrison G.P., Expansion planning of distribution networks considering uncertainties, Proceedings of the 44th International Universities Power Engineering Conference UPEC 2009, pp.1-5;
[20] Wang D.T., Ochoa L.F., Harrison G.P., Modified GA and data envelopment analysis for distribution network expansion planning under uncertainty, IEEE Transactions on Power Systems, vol.26, no.2, 2011, pp.897-904;
[21] Molga M., Test functions for optimization needs, http://www.zsd.ict.pwr.wroc.pl/files/docs/functions.pdf, 2005;
[22] Haupt R.L., Haupt S.E., Practical genetic alghorithms, 2nd Edition, John Wiley & Sons, 2004;
[23] Digalakis J.G., Margaritis K.G., An experimental study of benchmarking functions for genetic algorithms, IEEE International Conference on Systems, Management and Cybernetics, 2000 , vol.5, pp.3810-3815.

An Empirical Study from Industrial Design Engineering Students' Product Experiences with Intelligent Every Day Used Product

Emma Lógó
Budapest University of Technology and Economics
Faculty of Economic and Social Sciences
Institute of Applied Pedagogy and Psychology
Department of Ergonomics and Psychology,
Magyar Tudósok krt. 2. bldg Q1 , room A105. Hungary, H-1117
emma@erg.bme.hu

Ildikó Petruska
Budapest University of Technology and Economics
Faculty of Economic and Social Sciences
Institute of Business Sciences, Department of Management and Corporate Economics
Magyar Tudósok krt. 2. bldg Q1 , room B304. Hungary, H-1117
petruskai@mvt.bme.hu

Abstract: This paper is investigating the latent factors affecting Industrial Design Engineer students' everyday product use and analyse their product experiences. Different and distinct parts of product experience are frequently investigated, like usability, aesthetic judgments, or brand preferences. In this article authors examined a holistic aspect. Q-methodology is used for data collection and the sample is analyzed with a modified factor analysis. This method provides researchers a systematic and rigorously quantitative mathematical tool for examining human subjectivity. Q-set construction was the first step in order to reach the targeted aim. It was based on product experience case studies derived from students' everyday product interactions. Focusing on the group of intelligent everyday product was the next step. Q-sorts (the data collection) concentrated on this type of human-product interactions with 23 products chosen by Q-sorters. After the factor analysis a combination of experience structure was shown by means of 8 different factors.

Keywords: Q-methodology, Product experience, Industrial Design Engineer student, Intelligent product, Subjectivity

1. Introduction

Everyday experiences involve people who simply use and enjoy products. Industrial Design Engineer (IDE) students in Budapest University of Technology and Economics (BME) study designing new products and everyday product experiences. This also includes improving products that people use in everyday life. They need to learn about product construction and engineering, and of course material properties. In addition they are focused on social developments and design trends and are taught to apply their creativity in a systematic way. Students need to take into account what will happen to the product after being discarded that is why they are also concentrating on sustainability and environmental issues. Therefore IDE students meet every aspects of product experience during their studies.

This article focuses on IDE students' own product experiences using personal intelligent products in their everyday lives. According to Herring (1988), an intelligent product is both a product and a process and very difficult to define it. The product is the actual outcome of the process. The process, on the other hand, is a systematic way of producing something. Authors used the concept of intelligent product in this paper as a wide range of product meaning both software and hardware as well. One of the issues is that how creativity is displayed in IDE students' own intelligent product experiences. The other issue is the motivation of creating his or her product choices. The third issue is that what kind of feelings, or emotional relationships, or motivation factors can be detected about personal intelligent products in addition to creation and creativity. These questions are subjective and related to personal feelings.

Subjectivity, for product experience purpose, is defined simply as a person's point of view on personal importance of human-product interaction factors. Although it is clear that human actions are motivated on the basis of subjective perceptions and interpretations, the question is how and why exactly people see their world as they do. Following *Hekkert (2006)*, three components or levels of product experience have been separated: aesthetic pleasure, attribution of meaning, and emotional response. These levels depend on the humans' subjective opinions, judgments and feelings.

In this study we have to measure subjectivity and experience, and Q-methodology is an ideal tool to examine these types of research questions. Q-methodology is a relatively little-known form of research methodology within social and engineering science, even though it was improved more than 80 years ago (*Barry and Proops, 1999*). It is a qualitative approach based on mathematical statistical tools, which provides the study of subjectivity, a person's viewpoint, opinion, and attitude. In this paper Q-methodology was applied on IDE students' subjective feelings.

2. Theoretical background

2.1. Product Experience

Before starting the discussion on human-product interaction experience from the pont-of-view of intelligent products' perspectives it is important to summarize the theoretical background of product experience. Product experience always results from some interaction between user and product. This interaction is not necessarily restricted to instrumental or non-instrumental physical action, but many also consist of passive (often visual) perception, or even remembering or thinking of a product (*Desmet, Hekkert, 2007*). Human-product interaction and product experience is closely interwoven. *Figure 1*. provides a model of human-product interaction (*Hekkert, Schifferstein, 2008, 3*).

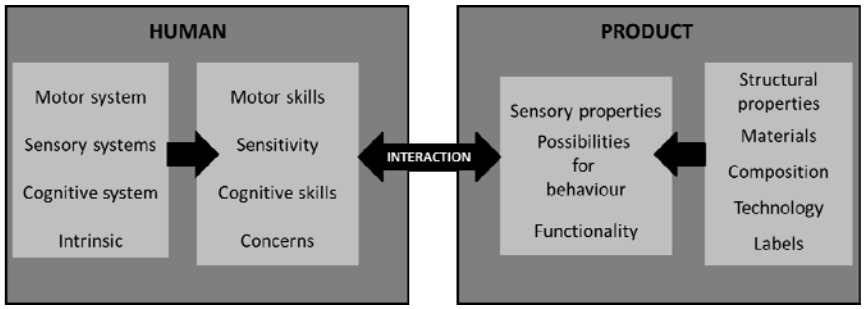


Figure 1
Modell of human-product interaction
Source: *Hekkert, Schifferstein, 2008, 3*).

Humans are interacting with their environment (and the products) through motor system, sensory systems, cognitive system and intrinsic constructs. Intelligent products have numberless connections with humans regarding product experience. They interact with humans on concrete levels:

- visual appearance (shape, colour, material, display),
- tactual experience (touch of controls, weight etc.),
- auditive experience (for example the sound of "click" or the ringtones etc.) and
- multisensory experience (e.g. playing with computer games or using navigation tool etc.)

On the other hand intelligent products interact with humans on more abstract, subjective, emotional "symbolic" levels, and they help to create the "owner's loyalty" like aesthetic experience (for example the subjective meaning of "beauty"), brand experience (like the producer's image, or the consumer's self-image when possessing the product), social level (the experience of belonging to a group), shopping experience, and the satisfaction with dealer service. It is clearly

visible that intelligent products' product experience is significantly more than the using experience itself.

Colours, shapes, sounds (or music) and words always have an emotional meaning. This emotional meaning is in part innate but also learned from our cultural and social environment. It is a question of design: design elements can only be successful if they hit the desired emotional spot. Cues, sensory, verbal and visual stimuli must find the best compromise among the level of product, the whole purchase and the consumption process. In this area it is increasingly difficult to compete with only the quality of product therefore new management solutions are formed focusing on emotional experiences targeting user groups. This new focus appears in emotional and symbolic product attributes emphasizing positioning, and the aim of product development is to impress all senses of the customer. In order to be successful on the market, the brand of a product must have a clear emotional place and message in the mind of the consumer (*Hausel, 2008*). This emotional brand essence results from the sum of all experiences coming from human-product interaction on both tangible and abstract levels. The subsequent implementation in marketing and brand communication is derived from the emotional brand positioning as a new and emerging field of marketing.

2.2. Q-methodology

2.2.1. Historical background

This presented work is on theoretical basis of the Q-methodology concept. Q-methodology is primarily an exploratory technique (*Watts and Stenner, 2005*). The idea behind the development of this methodology was to map the subjectivity of human mind. The examples of subjectivity are countless and include aesthetic judgment, appreciation of art, preferences for music, families' experiences after tragic events, and attitudes towards political groups. These were difficult - if not impossible - areas to be measured and reported scientifically by the conventional quantitative mathematical statistical methods available at 30's. Q-methodology emerged as a direct result of that deficiency. In the 1970s-1980s advanced computer programs appeared to perform mathematical statistical analysis of data derived by the Q methods. Authors have built up a model that deals with questions of environmental awareness and individual attitude.

Nowadays, Q-sorting has several benefits (*Thomas, Watson 2002*):

- Q-sort offers a means for an in-depth analysis of a small sample;
- It is useful in exploratory research;
- A well-founded theoretical background leads and assists its usage although the practical use is yet insignificant;
- It captures subjectivity in operation through a person's self-reference;
- Participants should be randomly selected, this helps to easily build samples;

- It may be administered over Internet; (but in this study we made personal interviews in a conventional way)
- Analysis techniques help protect respondent's self-reference from researcher influence.

Q methodology "*combines the strengths of both qualitative and quantitative research traditions*" (*Dennis and Goldberg, 1996:104*, *Sell and Brown, 1984*). As such, subjectivity is always anchored in self-reference, that is a person's internal frame of reference, and, Q studies from conception to completion adhere to the methodological axiom that subjectivity is always self-referent. (*McKeown and Thomas, 1988*).

2.2.2. Statistical background

Q analysis as a tool of mathematical statistical analysis typically includes sequential application of three statistical procedures: correlation analysis, factor analysis and the computation of factor scores. (*McKeown and Thomas, 1988*). Factor analysis is a statistical method of data reduction used to identify a small number of latent constructs (factors) that explain unobservable relationships among a large number of variables. The main applications of factor analytic techniques are:

(1) to reduce the number of variables; and

(2) to detect structure among variables, in order to classify or reduce the variables.

Therefore, factor analysis is applied as a data reduction or structure detection method. Firstly, Q-methodology inverts the factor extraction and correlates the person over a set of variables instead of opposite. The "conventional" or "usual" factor analysis is known as R-technique. (*Cattel, 1966 and Minke, 1997*) In this way, as it can be seen in the left part of Figure X the data matrix has variables as columns and subjects as rows. The data matrix factored in Q-technique as the right part of Figure 2 shows subjects as columns and variables as rows. In R-technique, more subjects (Persons) than variables are needed but in Q-technique, we need more variables than subjects *(Thompson, 1980).* The results of a Q-technique factor analysis differ from a usual typology in which each person fits one, and only one discrete category. Unless exceptionally simple structure is achieved with the factor analysis, each person may be related to more than one typological factor in Q-technique *(Gorsuch, 1983).*

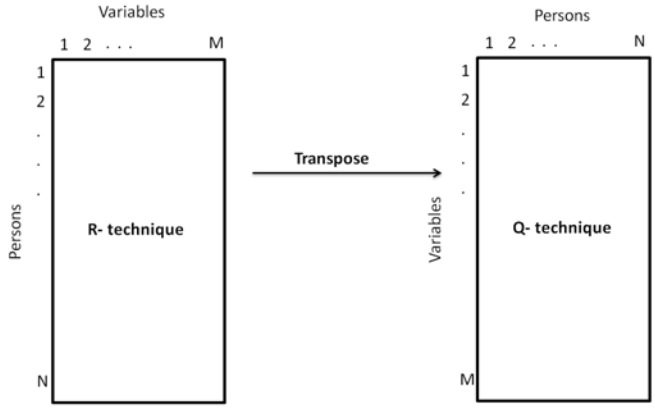


Figure 2:
The difference between R-technique and Q-technique data matrix
Source: own compilation

Secondly, Q-methodology is looking at the way the isolated factors – which are in the case of Q-methodology represent persons or more precisely their Q-Sorts – are rotated. Therefore, people of the same group or having the same factor will have a similar pattern of chosen statements. Q method is designed to understand the subjective expressions and viewpoints of participants and try to group them (*Watts and Stenner 2005*).

The population, in the conventional research methodological term, refers to the group of people in which the results of the study can be applied. The sample refers to those people on which the study is actually been conducted. Classical test theory assumes that each person has a "true score" (T) that would be obtained if there were no errors in measurement. A person's true score is defined as the expected number-correct score over an infinite number of independent administrations of the test. Unfortunately, test or questionnaires never observe a person's true score, only an observed score, X. It is assumed that observed score = true score plus some error:

$$X=T+E \quad (1)$$

where,

X: observed test score [-]

T: true test score [-]

E: error [-]

In Q methodology, the population and the sample is not as rigidly defined as in quantitative research and have no strict regulation in the connection between them. The sample needs not to be randomly drawn from the population. Often the persons are chosen for the research because they have special relevance or hold

strong views about the topics. Also the sample size can be relatively small. In fact, the subjective distortion (the "error") can be studied with Q methodology.

3. Objective and Research Method

3.1. Objective

The first reason to adopt the Q methodology in the field of product experience is that it allows the participants to express their subjectivity without confining them to the researcher's categories. Through a "mediate product", (in this study the mediate product will be IDE students' personal used intelligent accessories) and by applying Q-methodology, subjective experience category can be achieved without influencing them to the researcher's preliminary visions. The model of methodology can be seen in Figure 3:

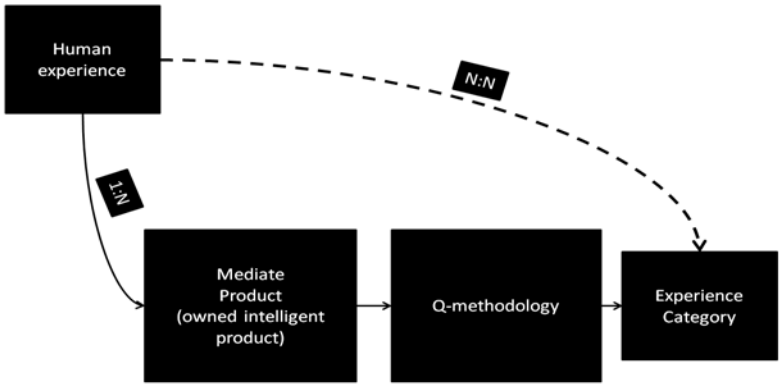


Figure 3:
Model of methodology
Source: own compilation

Table 1:
Selected objects by participants

Product	Q-sorter Gender	Q-sorter Age
Cannon MP620 printer	Female	23
Compaq laptop	Male	26

Dell Inspiron laptop	Female	24
Dell laptop	Female	25
Dicota PC mouse	Female	25
Farmerama flash game	Female	22
Fujitsu laptop	Female	23
GoogleCrome browser	Male	23
Igo8 Navon N47	Male	26
Innocentive.com	Male	23
IPhone 3G	Female	24
IPod nano 2G	Male	26
IPod nano 3G	Female	23
IPod touch 4G	Male	30
Istockphoto.com	Male	23
Logitech loudspeaker	Male	24
Mac Book laptop	Female	26
MSI laptop	Female	23
Nikon D90 camera	Male	30
Nokia 6020 telephone	Male	25
Nokia E51 telephone	Male	27
Prezi.com	Female	23
Sony Ericsson K550i telephone	Male	24

This research deals with intelligent products used every day and the related user attitudes are based on 23 objects of Q-methodology. Participation was voluntary, without any gift or payment. Objects were selected by participants (Q-sorters). IDE students were asked to pick a weekly or more often used personal intelligent product (software or hardware) as the object of Q-sorting. The participants' attitude to the selected product was not preconditioned. Only they first-hand experience and the personal use were important, and not their feelings or content of relationship with the product. *Table 1* shows the chosen products, and the participants' most important demographic data (gender and age).

3.2. Method

Having identified 'product experience' as an area of interest for the research, authors had to generate a series of statements on this topic. The significant source for statements was 10 "Product experience case studies", but statements were also used from academic literature (*Hekkert Schifferstein, 2008*). Intrinsic constructs have generated more than 150 possible statements. After some initial piloting, 52 statements were found to be reasonable, both for the Q-sorter and for the research aim as well.

Then the Q-set was disposed to Q-sorting. 23 Q-sort completed with selected objects. First the Q set was given to the respondent in the form of a pack of randomly numbered cards. Each card contained one of the statements from the Q set, and the number was on back of the card.

The respondent was instructed to rank the statements according to the basic rules of Q sorting, because the Q-sort distribution was forced in a way, that a certain number of items were prescribed for each rank. The subject was free, however there was a barrier (number of slots) to place an item anywhere within the distribution (*McKeownn and Thomas, 1988*). The score sheet was discrete, ranging from "most not-significant" to "most significant". Q-sorting began with pre-selection. Participants were asked to select the least relevant 10 statements, that got to the middle of the distribution with 0 score on scale. After that they had to order statements on a pre-prepared scale as Figure 4 demonstrates.

After Q-sorting, a very short follow-up interview was applied to capture the subjects' reasoning for ranking the various Q-samples in their unique way. The analysis of the Q sorts was the next step. It is a purely technical, objective procedure, with using factor analysis. Data analysis involves three proceedings applied in order: correlation analysis, factor analysis and computation of factor scores. To assist in the statistical analysis of Q data IBM SPSS Statistics 19 software was used. In the course of factor analysis Varimax rotation was used. The final step before we started to describe and interpret the factors was the calculation of the statements' factor scores in every given factor. The whole method's set development, data collection and data analysis process is summarized by Figure 5.

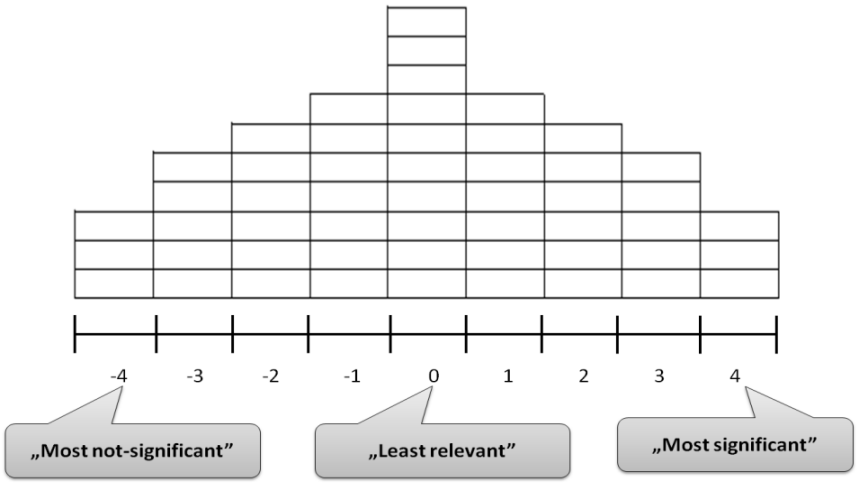


Figure 4:
Applied 9 point scale; and the taken quasi-normal formed distribution
Source: own compilation

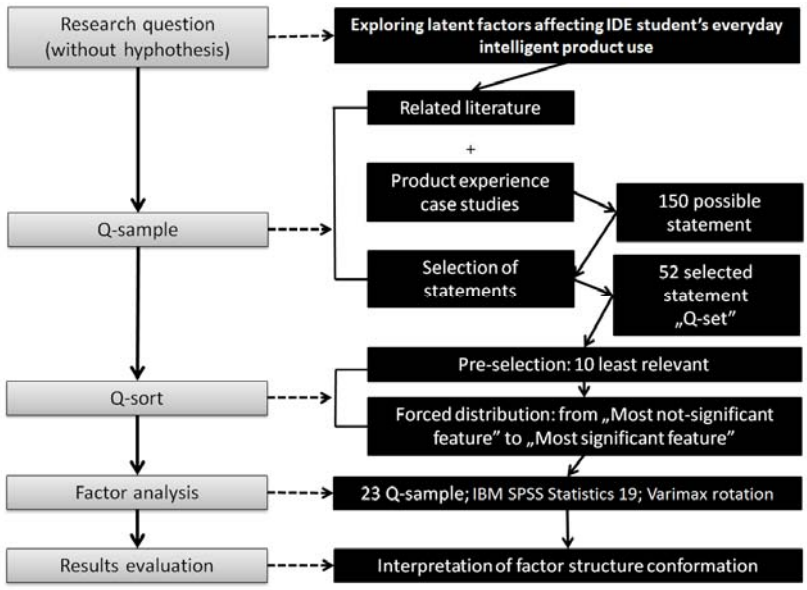


Figure 5
Set development, data collection and data analysis process
Source: own compilation

4. Results

First, the sample correlation matrix was investigated with Bartlett's test of sphericity. Bartlett's test is used to examine the hypothesis that the variables were uncorrelated in the sample. The result of the test was that each variable correlated perfectly with itself, but had no significant correlations with the other variables. On the other hand, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was calculated to examine the appropriateness of factor analysis. Our sampling's KMO measure of sampling adequacy was 0.576, this value (between 0.5 and 1) indicates that the factor analysis was appropriate.

The eigenvalue represents the total variance explained by each component (factor). Scree plot is a plot of the eigenvalues against the number of components (factors) in order of extraction. On this sample we explored 8 different factors with eigenvalue higher than 1. Figure 6 shows the scree plot and the resulting factors.
Rotated factor matrix (**Error! Reference source not found.**) contains the factor loadings of all the observed product experiences (variables) on all the factors extracted. The name of the obtained factors is based on the content of the factor. It derived from the factor scores which were calculated from every statement on each resulting factors. In this crucial interpretive step of the research, it is important that all statements participate in describing meaning to the factors obtained. By the highest and lowest factor scores for each statement characterized each factor.

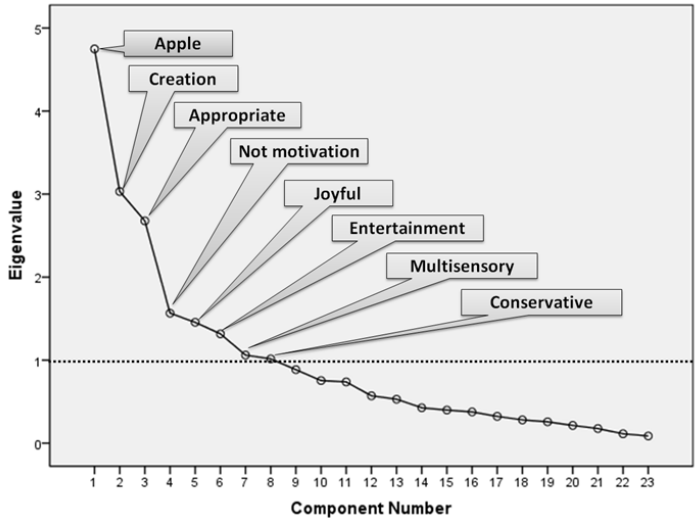


Figure 6
Scree Plot with the resulting 8 factor's eigenvalue
Source: SPSS 19

The first factor unexpectedly represents only **Apple** products. Pridefulness and positive feelings connected with the brand are the most significant experiences describe the factor and Apple products. It is very interesting that only Apple products belong to this factor, but from different product category, such as iPod nano media players and Mac Book laptop. The highest factor scores were related to the following Q-statements:

- Owning the product makes me feel proud.
- Using this product makes me relax.
- I feel that possessing this product refers to a standard of living of higher quality.
- The product in itself has an aesthetic appearance.
- Owning the product makes me feel special.
- The product is good against boredom.
- Possessing this product's brand means a lot to me.
- Using the product can easily become a regular habit.

The lowest factor scores are supported the highest scores. This product experience is the least of all annoying, and rather amusing than effective. These products are not compatible with other products in its category it is a crucial point that makes the owner feel special.

- Using the product is annoying.
- By using the product I can be more effective.
- It also contributes to the experience that this product is compatible with almost everything in its category.

As it can be seen on *Figure 6* and the second column of **Error! Reference source not found.**, the second factor comes out from some experience of **Creation**. Software products connected with work or other creations belong to this type of experience. The following statements describe the Creation factor:

- The product helps me to exploit my creativity.
- The product allows me to show how smart am I.
- It allows me to explore my abilities.
- The product allows me to manifest my fantasy.
- The product offers wide range of combinations.
- The product gives the experience of success.
- The product motivates me to use it.
- By using the product I can be more effective.

The lowest factor scores reveal that these products are exciting to use, and well-designed.

- Using the product is boring.
- The product design is boring.
- Using the product is annoying.

The third, fourth and fifth factors similarly belong to the experience of use, or experience of usability. These factors are different from the users' feelings which are related to the usability. The statements with lowest factor scores helped to

determine the difference among these three factors. The third factor was named **Appropriate**. Products used every day, such as printer, laptop and PC mouse, appertain to the factor of experience of "Appropriate". The 3. factor only refers to usability and does not contain creation content. The following statements had the highest factor scores:

- The usage of the product always has a goal.
- This product impresses me through sounds.
- The function and purpose of the product is recognizable by its appearance.
- The product can be used in private or in public environment.
- The product has ideal size.
- The product does not have any unnecessary functions.
- The product is easy to use.
- The product offers a wide range of combinations.

The lowest factor scores instance that this type of experience describes an easy to use product, but without creation:

- The product allows me to manifest my fantasy.
- Using the product is tiring.
- Using the product is annoying.

The fourth factor was named **Not motivation**. The statements with the highest factor scores have a very analogous content with the third factor:

- The usage of the product always has a goal.
- The product is easy to use.
- The product can be easily protected from damage.
- Using the product can easily become a regular habit.
- The product provides high level of experience for both sexes.
- The brand and the name are easy to recognize.
- The function and purpose of the product is recognizable by its appearance.
- By using the product I can be more effective.

Statements with the lowest factor scores allude to the real dissimilarity with the third factor. The third factor described more pleasant feelings than mentioned in the fourth factor. The most not-significant statements report the more unpleasant experiences, without happiness and motivation to use.

- Gives me self-confidence.
- Using the product makes me happy.
- The product motivates me to use it.

Table 2/A
Rotated factor matrix: Products and obtained factor loadings (1-4 factor)

	1. Apple	2. Creation	3. Appropriate	4. Not motivation
iPod touch 4G	**0,809**	-0,053	-0,069	0,154
iPhone 3G	**0,773**	0,029	0,029	-0,242
iPod nano 2G	**0,709**	0,181	0,234	-0,038
Mac Book laptop	**0,564**	-0,105	0,006	0,1
iPod nano 3G	**0,518**	-0,244	0,157	-0,382
Istockphoto.com	-0,003	**0,812**	-0,003	-0,104
Innocentive.com	0,004	**0,805**	0,021	0,053
Prezi.com	-0,055	**0,749**	0,184	-0,014
Dicota PC mouse	0,06	0,029	**0,83**	-0,098
Cannon MP620 printer	0,118	-0,021	**0,804**	0,164
Fujitsu laptop	-0,151	0,266	**0,471**	0,077
GoogleCrome browser	-0,241	-0,288	-0,037	**0,595**
MSI laptop	0,05	0,169	**0,574**	**0,583**
Dell laptop	0,183	0,371	-0,112	**0,564**
Dell Inspiron laptop	0,246	0,125	0,037	-0,004
Compaq laptop	0,231	0,36	0,174	-0,082
Farmerama flash game	0,204	-0,105	0,056	0,009
Logitech loudspeaker	0,189	-0,095	0,361	0,016
Igo8 Navon N47	0,14	0,223	0,076	-0,08
Sony Ericsson K550i telephone	0,024	-0,034	0,243	0,214
Nokia 6020 telephone	0,133	0,039	0,338	**0,446**
Nokia E51 telephone	-0,131	-0,088	0,242	0,036
Nikon D90 camera	**0,465**	0,086	-0,049	0,035

The fifth factor is, to some extent, similar to the third factor. It was named **Joyful**, and it belongs to everyday product usability, without special experiences, but with more significant amusement than the third and fourth factor. The fifth factor contains more positive feelings related to the brand and the brand evoked emotions.

- By using the product I can be more effective.
- Owning the product makes me feel proud.
- It also contributes to the experience that this product is compatible with almost everything in its category.
- The product in itself has an aesthetic appearance.
- The brand and the name are easy to recognize.
- Using the product can easily become a regular habit.
- The product helps me to show how I feel about myself.
- The product allows me to manifest my fantasy.

The most not-significant statements refer to the facts that it was not a unique and special experience with products, but the usage is not boring:

- Using the product is boring.
- I feel that possessing this product refers to a standard of living of higher quality.
- If I have problems with the use, help and guarantee is offered by the Producer.

The sixth factor adds **Entertainmen**t. Two products represent this experience. The flash game speaks for itself, but the post interviews elicited that the loudspeaker was a tool that had been used more for leisure time activities. The highest factor scores were related to the following Q-statements:

- The product is good against boredom.
- The function of the product can be easily understood.
- The product is easy to use.
- Using the product makes me happy.
- The product allows me to manifest my fantasy.
- The product provides high level of experience for both sexes.
- The product design makes me happy.
- The product never disappoints me.

Statements with the lowest factor scores confirmed the spare-time experiences, usage without goal and special emotions:

- Owning the product makes me feel special.
- The usage of the product always has a goal.
- Possessing the product means a lot to me.

Multisensory experiences belong to the seventh factor. These experiences based on touching and hearing and seeing, integrated with some brand connected emotional feelings. The following statements had the highest factor scores:

- This product impresses me through sounds.

- By using the product I can be more effective.
- It is a multimodal product experience.
- The use of the product is regulated in some aspects.
- This product impresses me through touch.
- The design and the quality of the available accessories go together with the product.
- I feel that possessing this product refers to a standard of living of higher quality.
- Using the product can easily become a regular habit.

Statements with the lowest factor scores reveal the products' weaknesses. Experiences belong to typical cell phone usage, because mobile phones almost always contain unnecessary extra functions. In this case these products are not status symbols, but only a product used every day without elevated emotions.

- Owning the product makes me feel proud.
- The product provides high level of experience for both sexes.
- The product does not have any unnecessary functions.

The last factor is named **Conservative**. This factor contains very mixed feelings. These are trustworthy classical products without inordinate emotional feelings. The brands of these products are important in this particular case. Nokia and Nikon is very similar; they are long-standing, reliable, but not typically "love brands", like Apple in the first factor. The highest factor scores were related to the following diverse Q-statements:

- The product design is boring.
- The product does not have any unnecessary functions.
- The product requires using my logical sense.
- The product has everlasting design.
- The product is good against boredom.
- The brand and the name are easy to recognize.
- The product can be used in private or in public environment.
- The function and purpose of the product is recognizable by its appearance.

Statements with the lowest factor scores confirmed the conservative experiences, usage without excitement:

- The product is always in fashion.
- It also contributes to the experience that the product is unique.
- The product design makes me happy.

Table 3/B
Rotated factor matrix: Products and obtained factor loadings

	5. Joyful	6. Entertainment	7. Multi-sensory	8. Conservative
iPod touch 4G	0,16	0,209	-0,147	0,067
iPhone 3G	0,246	0,037	0,103	-0,026
iPod nano 2G	-0,082	0,152	0,329	-0,039
Mac Book laptop	0,349	-0,189	0,056	-0,487
iPod nano 3G	0,206	-0,001	0,197	-0,074
Istockphoto.com	0,035	0,234	0,061	-0,046
Innocentive.com	0,063	-0,276	0,184	0,115
Prezi.com	0,315	-0,292	-0,04	-0,211
Dicota PC mouse	0,186	0,075	-0,011	0,083
Cannon MP620 printer	0,019	0,034	0,167	0,208
Fujitsu laptop	0,018	0,404	0,166	0,095
GoogleCrome browser	0,083	0,368	0,288	0,076
MSI laptop	-0,045	-0,027	0,116	-0,253
Dell laptop	0,277	0,159	0,109	0,34
Dell Inspiron laptop	**0,807**	0,046	0,008	0,026
Compaq laptop	**0,623**	0,029	0,251	0,07
Farmerama flash game	0,025	**0,776**	0,08	-0,054
Logitech loudspeaker	0,278	**0,588**	-0,108	0,453
Igo8 Navon N47	0,139	0,019	**0,852**	-0,023
Sony Ericsson K550i telephone	**0,454**	0,215	**0,493**	0,145
Nokia 6020 telephone	-0,096	0,101	**0,476**	0,353
Nokia E51 telephone	0,085	0,005	0,133	**0,778**
Nikon D90 camera	0,083	-0,399	-0,317	**0,496**

There are 4 products in the sample which cannot be clearly classified only one factor in the structure. MSI laptop belongs to the "Not motivation" factor (factor load: 0,583), but in "Appropriate" factor has hardly differed from factor load, 0,574. This means that the product relates to both experiences. The content of MSI laptops' Q-sort equally participated in both factors.
Sony Ericsson K550i telephone and Nokia 6020 telephone mainly belongs to the "Multisensory" experience (factor loads: 0,493 and 0,476). But Sony Ericsson K550i telephone is only slightly less factor load in the "Joyful" factor (0,454). According to follow-up interview, in this case the subject uses his telephone to listening to music, taking photographs too, and likes the brand. The owner of Nokia 6020 telephone said in the follow-up interview that he wanted a new mobile phone soon. The products belong primarily to the "Multisensory" factor still because of multimodal experiences.
Nikon D90 camera primarily belongs to the "Conservative" factor, with 0,496 factor load, but in "Apple" factor has similarly high factor load, 0,465. The content of the Q-sort equally participated in both factors. The follow-up interview revealed that the owning of Nikon D90 made the Q-sorter feel proud, like the Apple products.

5. Conclusion

The factor structure is not based on products category. The basis was the user's experience. For example the laptops can prove this. In the research five different laptops with four different brands were examined. After the analysis, in the formed factor structure, the laptops belong to four different factors because of the different related experiences. Obviously, MacBook laptop belongs to **"Apple"** experience. Fujitsu laptop is an **"Appropriate"** product to the owner. MSI laptop is a mixed experience between **"Appropriate"** and **"Not motivation"** category. In the research we had two Dell laptops. Both of them are owned by female users. For the first user her Dell laptop has **"no motivation"** to use. According to the follow-up interview this user owns her laptop for 2 years. The other female with the Dell Inspiron laptop felt **"Joyful"** to the using experience. It turned out from the follow-up interview, she had a new laptop for 2 months, and she just enjoyed the experience of the discovery.

The factor structure is not based on products' brand, except for Apple. In addition Dell laptops and two Nokia telephones were examined. The first experience with Nokia 6020 telephone belongs to the **"Multisensory"** factor. The other product, Nokia E51 telephone was with **"Conservative"** factor.

Creation is very important when somebody wants to be a designer. In this study the experience of creation appeared as a separated factor.
The most interesting result of the research was the "**Apple**" factor. The brand and the company are in accordance with the needs of the IDE students. Apple never published a mission statement, but the main milestones could be to "make great

products", "constantly focusing on innovation", "believing in the simple, not the complex", "need to own and control the primary technologies behind the products". These buzzwords are coincidence with what IDE students learned, for this reason these products has special place in minds.

6. Ackowledgements

This work is connected to the "Product Experience Case Study 2010" competition. These competitions are supported by the *Heller Farkas Scholarship* BME.

References

[1] Barry, J. & Proops:'Seeking Sustainability Discourses with Q Methodology, Ecological Economics, 1999. 28, pp.337-345

[2] Barry, J., & Proops, J.: Seeking sustainability discourses with Q methodology, Ecological Economics, 1999. 28, 337-345.

[3] Brown, S. (1980). Political subjectivity: Applications of Q methodology in political science. New Haven, CT: Yale University Press.

[4] Cattell, R. B.: The data box: Its ordering of total resources in terms of possible relational systems. In R. B. Cattell (ed.), Handbook of multivariate experimental psychology; Chicago: Rand McNally. 1966. pp. 67-128.

[5] Dennis, K.E., & Goldberg, A.P.: Weight control self-efficacy types and transitions affect weight-loss outcomes in obese women. Addictive Behaviors, 1996. 21, 103-116.

[6] Desmet, P.M.A., & Hekkert, P.: Framework of product experience, International Journal of Design, 2007. 1(1), 57-66.

[7] Gorsuch, R. L.: Factor Analysis (2nd ed.). Hillsdale, NJ: Erlbaum. 1983.

[8] Häusel, G. F.:Brain View: Warum Kunden kaufen, Haufe MediengruppeRudolf Haufe Verlag GmbH & Co. KG, Niederlassung Planegg/München, 2008.

[9] Hekkert, P., & Schifferstein, H.: Product experience, Amsterdam, Elsevier Science 2008. 1 ed., pp. 3.

[10] Hekkert, P.: Design aesthetics: Principles of pleasure in product design. Psychology Science, 2006. 48(2), 157-172.

[11] Herring, J. P.: Building a Business Intelligence System, Journal of Business Strategy, 1998. Vol. 9:3

[12] McKeown, B., & Thomas, D.: Q methodology. Sullivan, J. L., & Niemi, R. G. (Eds.): Quantitative applications in the social sciences.Newbury Park: Sage Publications. 1988.

[13] Minke, A.: The Six Two-Mode Factor Analytic Models; Annual meeting of the Southwest Educational Research Association, Texas: Austin, January, 1997.

[14] Robbins, P. & Krueger, R.: Beyond bias? The promise and limits of Q-method in human geography. Professional Geographer 200. 52(4):636–648.

[15] Sell, D.K., & Brown, S.R.: Q methodology as a bridge between qualitative and quantitative research: Application to the analysis of attitude change in foreign study program participants. In J.L. Vacca & H.A. Johnson (Eds.), Qualitative research in education (Graduate School of Education Monograph Series) (pp. 79-87). Kent, OH: Kent State University, Bureau of Educational Research and Service. 1984.

[16] Thomas, D. & Baas, L.: The issue of generalization in Q Methodology: "Reliable schematics" revisited, Operant Subjectivity, 1992, 16(1), pp. 18-36.

[17] Thomas, D. M. & Watson, R. T.: Q-Sorting and MIS Research: A Primer, Communications of the Association for Information Systems 2002. 8, 141-156.

[18] Thompson, B.: Validity of an evaluator typology. Educational Evaluation and Policy Analysis, 1980. 2, 59-65.

[19] Watts, S., & Stenner, P.: Doing Q methodology: Theory, method and interpretation, Qualitative Research in Psychology, 2005. 2, 67–91.

Alpha-Numeric Notation for one Data Structure in Software Engineering

Sead Mašović

Computer Science Department, Faculty of Science and Mathematics, University of Niš, P.O. Box 224, Višegradska 33, 18000 Niš, Serbia
sead.masovic@pmf.edu.rs

Muzafer Saračević

Computer Science Department, Faculty of Science and Mathematics, University of Niš, P.O. Box 224, Višegradska 33, 18000 Niš, Serbia
muzafers@uninp.edu.rs

Predrag Stanimirović

Computer Science Department, Faculty of Science and Mathematics, University of Niš, P.O. Box 224, Višegradska 33, 18000 Niš, Serbia
pecko@pmf.ni.ac.rs

Abstract: This paper presents one way to store balanced parentheses notations in shortened form in order to lower memory usage. Balanced parentheses strings are one of the most important of the many discrete structures. We propose new method in software engineering for storing strings of balanced parentheses in shortened form as Alpha-numeric (AN) notation. In addition, our algorithm allows a simple reconstruction of the original strings. Another advantage of the presented method is reflected in savings of working memory when it comes to deal with combinatorial problems. The proposed method is implemented in Java environment.

Keywords: Balanced Parentheses, Catalan number, Binary tree, Software engineering, Java programming.

1 Introduction and Preliminaries

Data structure and software engineering are an integral part of computer science. A binary tree is a tree of data in which each node has at most two descendant nodes, usually distinguished as "left" and "right". Many powerful algorithms in computer sciences and software engineering are tree based algorithms.

The most renown representation of trees is the balanced parentheses [4,8,10,11]. The tree is represented by a string *P* of balanced parentheses of length 2*n*. A node is represented by a pair of matching parentheses "()" and all sub-trees rooted at the node are encoded in an order between the matching parentheses [2,13].

Suppose you would like to form valid sequences of *n* pairs of parentheses, where a string of parentheses is *valid* if it contains an equal number of open and closed parentheses and each open parenthesis has a matching closed parenthesis. For example, "(()())" is valid, but "())()(" is not. A string of parentheses is valid if there is an equal number of open and closed parentheses. The most simple method for presentation of balanced parentheses is to use Bit-string, which uses bit 1 to represent "(" and to use bit 0 to represent ")" in order to represent a well formed Balanced Parentheses (BP shortly).

For example, the Bit-string of the BP expression ()((()))(()) is given by 101110001100.

The number of different well formed parentheses strings represented by Bit-strings $b_{2n}\, b_{2n-1} \ldots b_1$ is given by the Catalan number [3, 6].

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!n!} \tag{1}$$

For example, for n=3, we have

$$C_3 = \frac{(2\times 3)!}{(3+1)!\times 3!} = \frac{6!}{4!\times 3!} = 5$$

and there are five sequences of six balanced parentheses (three pairs of parentheses) as presented in Figure 1.

((())) ()()() (())() ()(()) (()())

Figure 1
Different valid combinations of balanced parentheses for C_3

The BP representation is obtained by traversing the tree in depth-first order writing an open parenthesis when a node in first encountered and a closing parenthesis when the same node is encountered again while going up after traversing its sub-tree. The first representation of trees was proposed in 1989 by Jacobson [5]. The Jacobson representation is called *Level Order Unary Degree Sequence* (LOUDS), which lists the nodes in a level-order traversal. An alternative tree representation based of strings of open and closing parentheses is proposed ten years after results that gave Jacobson et al. [9,10].

In [1] Benot introduced *Depth First Unary Degree Sequence* (DFUDS) representation of an *n*-node tree. DFUDS combines the LOUDS and BP

representations. Three different ways for representing a tree (LOUDS, DFUDS and BP), are presented on Figure 2.

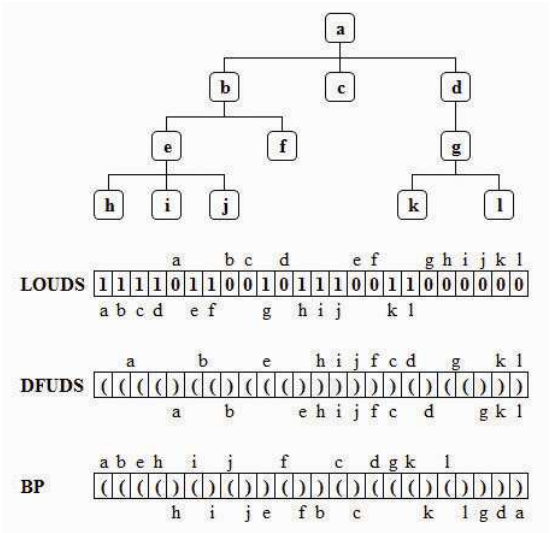


Figure 2
Different representations of tree

To minimize the memory space requirements, we use a bit-string as the basis for introducing further rules for shortened processes.

2 Transformation from BP to AN Notation

In this section we present the way how we get *Alpha-numeric* (shortly AN) notation through shortening process in order to obtain a large reduction of storage space. The shortening procedure is described in the Algorithm 1.

Algorithm 1. Transformation from BP to AN

INPUT : BP notation
m = BP.lenght
Replacement
for (i = 1; $i \leq m$; i++)
(→ 1 AND) → 0
Output1 = BP1 // *Binary Equivalent*
Elimination
Delete first element (1) and last element (0) in BP1
Output2 = BP2 // *Short Binary Eq.*
Selection

```
    e = BP2.element[i];
    m1 = BP2.lenght;
if (Case 2 - binary pairs){
    First binary pair= e[i] for i={1, 2}
    Last binary pair= e[i] for i={m1-1; m1}
    RestBP2= BP2 without First and Last binary pair
    replace (First AND Last binary pair → Alpha2)          // based on Table A3
    Output3 = Alpha2 + RestBP2                             // Alpha Binary
}
if (Case 3 - bit binary group){
    First three bit binary group= e[i] for i={1,2,3}
    Last three bit binary group= e[i] for i={m1-2; m1-1; m1}
    RestBP3= BP2 without First and Last three bit binary group
    replace (First AND Last three bit binary group → Alpha3)   // based on Table A4
    Output3 = Alpha3 + RestBP3                             // Alpha Binary
}
Conversion
if (Case 2 - binary pairs){
    RestBP2 → DecimalEq2.
    Output4 = Alpha2 + DecimalEq2.                         // Alpha Decimal
}
if (Case 3 - bit binary group){
    RestBP3 → DecimalEq3.
    Output4 = Alpha3 + DecimalEq3.                         // Alpha Decimal
}
OUTPUT : AN notation
```

Algorithm 1 consists of four phases, called *replacement, elimination, selection and conversation*):

Step 1 (replacement): Form a binary equivalent of given BP notation, called *b-string*. A bit-string uses bit `1` to represent "`(`" and use bit `0` to represent "`)`".

Step 2 (elimination): The *First shortened form* is obtained by omitting the first and the last bit from the b-string notation. The correctness of this elimination is ensured by the rule that every BP representation begins with the open parenthesis and ends with the closed parenthesis.

Step 3 (selection):

Case 2 - binary pairs*: The *Second shortened form* is obtained by grouping the first two and the last two binary digits of the *First shortened form*. Then the Alpha notation record is based on table grouping (Table A3 - Appendix). The central part (if any) is prescribed.

Case 3 - bit binary group*: In this case, *Second shortened form* is obtained by grouping the first three and the last three binary digits. Then the Alpha notation

record is derived using the special table grouping (Table A4 - Appendix). The central part (if any) is prescribed.

Step 4 (conversation): The final form of the AN notation is obtained by prescribing alpha entry from the second shortened form in the table and converting its central part (the rest) into a decimal number.

The final form of the AN notation is obtained by prescribing alpha entry from the second shortened form in the table and converting its central part (the rest) into a decimal number. Let us mention that Table A1 (Case 2 - binary pairs) and Table A2 (Case 3 - bit binary group) are presented in appendix. Also, the shortening process for the actual value of *n* is illustrated from the initial balanced parentheses over the binary equivalent and usage of special tables to get the AN record as short as possible.

Figure 3 presents the shortening process for one case from the Table A1.

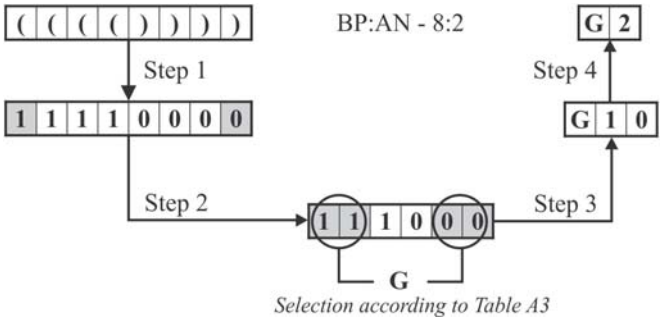


Figure 3
Conversion process of BP notation to AN notation

3 Reverse Transformation from AN to BP Notation

Reverse transformation from AN to BP notation. Based on Algorithm 1 we have possibility to define the reverse transformation of AN notation to the original form of BP notation. The reverse transformation consists of four phases. In the case of the reverse process phase of elimination becomes phase of addition. Description of reverse transformation from AN to BN is given below.
Figure 4 illustrates the reverse construction for the groups of two bits, how to get the original form of BP notation. The same process is taken when it comes for grouping of the initial three bits with ending three bits. The only difference occurs in the selection phase. The conversion in this phase is based on the usage of Table

A3 or Table A4 given in Appendix. More precisely, Table A3 is used for the reverse transformation of the group of two bits, while he reverse transformation of the group of three is based on the usage of Table A4.

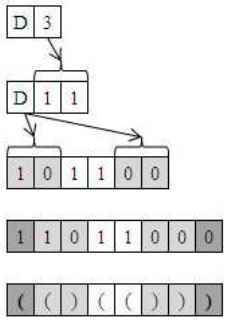


Figure 4
Process of reverse construction

Step 1 (Conversation): Decimal number `3` is converted to binary equivalent. This binary equivalent represents the central part of generated BP notation.

Step 2 (Selection): Alpha notation `D` transforms into corresponding two pairs of binary digits on the basis of rules from the Table A3. The first pair is placed in the beginning and the second at the end of the actual record.

Step 3 (Addition): Add bit `1` at the beginning of string obtained in Step 2 and add the bit `0` at the end of this string.

Step 4 (Replacement): Replace the bit `1` by the open parenthesis "`(`" and the bit `0` by the closing parenthesis "`)`" and so we get the original form of BP notation.

4 Some Application of the Presented Method

From the aspect of storage, it is very important to make good choices which relates to the application of the most appropriate data structure. A binary tree is an important type of structure which occurs very often in computer science.

BP notation is appropriate representation of binary trees. Balanced parenthesis can be applied in representation of Catalan numbers, which is the basis of many combinatorial problems. Another important application of BP notation is in the process of recording and storing polygon triangulation. This procedure is crucial in computer geometry and graphics in 3D view of images. Increasing the number of polygon vertices drastically increases the number of possible convex polygon triangulations. In order to reduce the memory space requirements, in our paper [12] we propose a shortened form (similar to AN notation) for the storage of generated triangulations. This shortened form presents a unique key for each graph or any combination of triangulations for convex polygons. Another way for representing the polygon triangulation is usage of the Reverse Polish notation, This approach is presented in the paper [7]. Compared to the record that provides Reverse Polish notation, presented AN notation gives much shorter record for the same triangulation. For example, to record one triangulation of hexagon takes eight bits in Reverse Polish notation, while AN notation takes one character and one integer or three bits.

In general case, presented AN method can serve as a model for shortening process in all other problems which are based on Catalan numbers [6]. For example, some of these problems are *Correctly parenthesized expression, Binary records from Lukasiewicz's algorithm, Ballot problem, Problem of the lattice path* and etc. The AN method should be used to save memory space in the storage of records.

5 Implementation Details and Experimental Results

Presented method is implemented in *Java NET Beans environment*. The advantages of Java over the other programming languages are numerous. First of all, programming in the Java programming language is one of the highest degree of programming. These written programs are easily portable from one platform to another.

Our application follows all the steps in obtaining Alpha-numeric notation which is given in Algorithm 1. In an Appendix we gave part of *Java source code*, which is responsible for calling the rules of Alpha notation for groups of binary pairs or bit binary group in order to get the shortened form. Figure 5 present part of the application that implements Algorithm 1 with additional options for presenting and storing results.

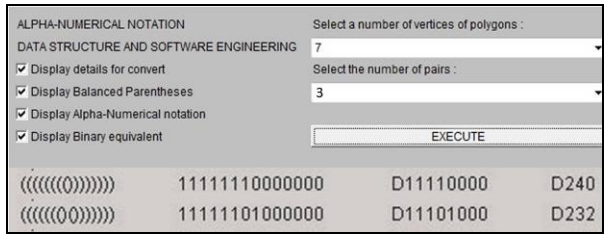
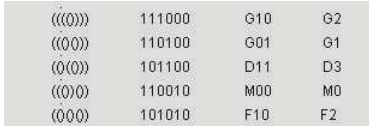
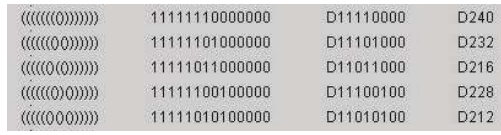


Figure 5
Java application

Figure 6 presents result of the application for both case (binary pairs and bit binary group).



Result screen of the execution of applications for *i=2* and *n=4*



Result screen of the execution of applications for *i=3* and *n=8*

Figure 6
Java panel for both case (binary pairs and bit binary group)

Based on the Algorithm 1 we can set ratio (R) of input (BP notation) and output (AN notation) with the equations:

$$R = \frac{BP}{AN} \tag{2}$$

Where is $BP = 2n$ and $AN = 2(n - i - 1)$ while n is index of Catalan number (C_n) and i is the number of pairs who can take the value from the set {2,3}.

Based on equation (2) in Table 1 are presented ratio of BP and AN notations from two aspects: ratio in the number of characters and ratio in the size of the output file in which the results are stored.

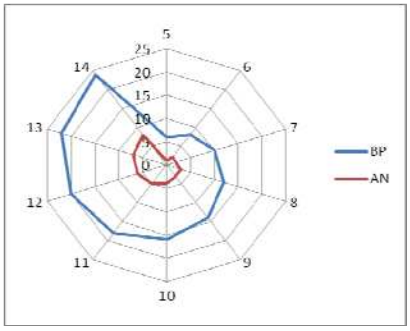
Table 1
Experimental results for testing application

n	**Number of characters**			**File size (KB)**		
	BP	**AN**	***R***	**BP**	**AN**	***R***
5	6	1	6.00	0.5	0.3	1.67
6	8	2	4.00	1	0.6	1.67
7	10	2	5.00	2	1	2.00
8	12	3	4.00	3	1.4	2.14
9	14	3	4.67	7	3	2.33
10	16	4	4.00	26	9	2.89
11	18	5	3.60	95	31	3.06
12	20	6	3.33	386	122	3.16
13	22	7	3.14	1378	446	3.08
14	24	8	3.00	4792	987	4.85

PC performance for testing results: *CPU – Intel (R) Core (TM) 2 Duo, T7700, 2.40 GHz, L2 Cache 4 MB (On-Die, ATC, Full-Speed), RAM Memory - 2 Gb, Graphic card - NVIDIA GeForce 8600M GS.*

Graph 1 presents ratio of shortening process in the number of characters applying AN notation compared to BP notation. Testing was done for n={5,6,…,14}.

Graph 1
Shortening process in the characters



Based on the testing results the advantage of using *AN* notation is when it come to the larger value for *n*, which can be seen in the graph 1. From another point of view if we look at the size of the output file, we can see that with increasing value of *n* is dramatically increasing shortening ratio of the output file for the AN notation (for example, *n*=14, for BP notation output file is 4792 kb while for AN notation is 987kb, which is almoust five times less). For all testing results specified in Table 1 for *n*={5,..,14} we get average shortening of 2,69 time less.

Conclusion

We develop an algorithm that can convert Balanced Parentheses notation in a new shortened Alpha-Numeric notation. With this work we present a way how to reduce BP notation using the appropriate rules in order to lower memory usage in storing. Advantages of presented method is reflected in savings of working memory in the process of testing implementations and thus achieve much better results in terms of speed of execution. Another advantage is that the stored values can be converted through reverse construction to get initial BP notation. We provide an evaluation of the algorithm on a Java implementation.

AN notation may find their application mainly in the problem of triangulation of a convex polygon . It is important to note that the application of this method of recording results may find usage in some combinatorial problems which are based on Catalan numbers.

References

[1] Benoit, D., Demaine, E. D., Munro, J. I., Raman, R., Raman, V., Rao S. S. Representing Trees of Higher Degree, Algorithmica, 2005, Vol. 43, No. 4, pp. 275-292.

[2] Evans D. J., Abdollahzadeh, F. Ecient Construction of Balanced Binary Trees, The Computer Journal, 1983, Vol. 26, No. 3, pp. 193-195.

[3] Geary, R. F., Rahman, N., Raman, R., Raman, V., A Simple Optimal Representation for Balanced Parentheses, Theoretical Computer Science, 2006, Vol. 368, No.3, pp. 231-246.

[4] Gog, S., Fischer, J. Advantages of Shared Data Structures for Sequences of Balanced Parentheses, DCC'10 Proceed. Data Compression Conf. 2010, pp. 406-415.

[5] Jacobson, G. Space-efficient static trees and graphs, In Proceedings of the 30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, North Carolina, IEEE, 1989, pp. 549–554.

[6] Koshy, T. Catalan Numbers with Applications, Oxford University Press, New York, 2009.

[7] Krtolica, P.V., Stanimirovic, P.S., Stanojević, R. Reverse Polish notation method in constructing the algorithms for polygon triangulation, Filomat, 2001, Vol. 15, pp.25–33.

[8] Lu, H., Yeh, C. Balanced Parentheses Strike Back, ACM Transactions on Algorithms (TALG), 2008, Vol. 4, No.3, pp. 1-13.

[9] Munro, I., Raman, V. Succinct Representation of Balanced Parentheses and Static Trees, SIAM Journal on Computing, 2001, Vol. 31, No.3, pp. 762-776

[10] Munro, I., Raman, V., Succinct representation of Balanced Parentheses, static trees and planar graphs, Proceedings of the 38th Annual Symposium on Foundations of Computer Science, IEEE, Miami Beach, Florida, 1997, pp. 118–126.

[11] Ruskey, F., Williams, A. Generating balanced parentheses and binary trees by prefix shifts, In CATS '08: Fourteenth Computing: The Australasian Theory Symposium, Vol. 77 of CRPIT, Wollongong, Australia, 2008.

[12] Saračević, M., Stanimirović, P., Mašović, S., et al., Implementation of some algorithms in computer graphics in Java, TTEM - Technics Technologies Education Management, 2013, Vol. 8, No.1, pp. 293-300.

[13] Tsay, J. Designing a systolic algorithm for generatng well-formed parenthesis strings, Parallel Process. Lett. 2004, Vol. 14, pp.83-97.

Appendix

Table A1
The process of converting from BP to AN for case of binary pair, n={1,2,...4}

n	Number of combinations for Catalan number of n	Balanced parentheses	Binary equivalent	First shortened form	Second shortened form	Final form
1	1	( )	1 0			0
2	1	( ) ( )	1 0 1 0	0 1		1
	2	(( ))	1 1 0 0	1 0		2
3	1	((( )))	1 1 1 0 0 0	1 1 0 0		G
	2	(( ) ( ))	1 1 0 1 0 0	1 0 1 0		F
	3	(( )) ( )	1 1 0 0 1 0	1 0 0 1		E
	4	( ) (( ))	1 0 1 1 0 0	0 1 1 0		C
	5	( ) ( ) ( )	1 0 1 0 1 0	0 1 0 1		B
4	1	(((( ))))	1 1 1 1 0 0 0 0	1 1 1 0 0 0	G 1 0	G 2
	2	((( ) ( )))	1 1 1 0 1 0 0 0	1 1 0 1 0 0	G 0 1	G 1
	3	(( ) (( )))	1 1 0 1 1 0 0 0	1 0 1 1 0 0	D 1 1	D 3
	4	((( )) ( ))	1 1 1 0 0 1 0 0	1 1 0 0 1 0	M 0 0	M 0
	5	(( ) ( ) ( ))	1 1 0 1 0 1 0 0	1 0 1 0 1 0	F 1 0	F 2
	6	( ) ((( )))	1 0 1 1 1 0 0 0	0 1 1 1 0 0	A 1 1	A 3
	7	( ) (( ) ( ))	1 0 1 1 0 1 0 0	0 1 1 0 1 0	C 1 0	C 2
	8	(( )) (( ))	1 1 0 0 1 1 0 0	1 0 0 1 1 0	F 0 1	F 1
	9	( ) ( ) (( ))	1 0 1 0 1 1 0 0	0 1 0 1 1 0	C 0 1	C 1
	10	((( ))) ( )	1 1 1 0 0 0 1 0	1 1 0 0 0 1	H 0 0	H 0
	11	(( ) ( )) ( )	1 1 0 1 0 0 1 0	1 0 1 0 0 1	E 1 0	E 2
	12	( ) (( )) ( )	1 0 1 1 0 0 1 0	0 1 1 0 0 1	B 1 0	B 2
	13	(( )) ( ) ( )	1 1 0 0 1 0 1 0	1 0 0 1 0 1	E 0 1	E 1
	14	( ) ( ) ( ) ( )	1 0 1 0 1 0 1 0	0 1 0 1 0 1	B 0 1	B 1

Table A2
The process of converting from BP to AN for case of bit binary group, n={6,7,8,9}

n	Number of combinations for Catalan number of n	Balanced parentheses (several examples)	Binary equivalent	First shortened form	Second shortened form	Final form
6	132	((((((())))))) (((((()())))))	1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 0 1 0 0 0 0 0	1 1 1 1 1 0 0 0 0 0 1 1 1 1 0 1 0 1 0 0 0	D 1 1 0 0 D 1 0 1 0	D 1 2 D 1 0
7	429	(((((((()))))))) (((((()()))))) (((((()(())))))	1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 0 0 0 0 0 0 1 1 1 1 1 0 1 1 0 0 0 0 0 0	1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 0 1 0 0 0 0 0 1 1 1 1 0 1 1 0 0 0 0 0	D 1 1 1 0 0 0 D 1 1 0 1 0 0 D 1 0 1 1 0 0	D 5 6 D 5 2 D 4 4
8	1430	((((((((())))))))) (((((((()()))))))) (((((()(())))))) (((((()()()))))) ((((()))))(())	1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 1 1 1 1 1 0 1 0 1 0 0 0 0 0 0 0 1 1 1 1 0 1 1 1 0 0 0 0 0 0 0 0	1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 0 0 0 0 0 1 1 1 1 1 0 1 1 0 0 0 0 0 0 1 1 1 1 1 0 1 0 1 0 0 0 0 0 1 1 1 1 0 1 1 1 0 0 0 0 0 0	D 1 1 1 1 1 0 0 0 0 D 1 1 1 1 1 0 1 0 0 0 D 1 1 1 1 0 1 1 0 0 0 D 1 1 1 1 0 1 0 1 0 0 D 1 1 1 0 1 1 1 0 0 0	D 2 4 0 D 2 3 2 D 2 1 6 D 2 1 0 D 1 4
9	4862	(((((((((()))))))))) ((((((((())))))))) (((((((()(())))))))	1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0	1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0	D 1 1 1 1 1 1 0 0 0 0 0 D 1 1 1 1 1 1 0 1 0 0 0 0 D 1 1 1 1 0 1 1 0 0 0 0 0	D 9 9 2 D 9 7 6 D 9 4 4

Table A3
Codebook for binary pair grouping

Number of combination	First binary pair		Alpha notation	Last binary pair	
	1	2		m - 1	m
1	0	1	A	0	0
2	0	1	B	0	1
3	0	1	C	1	0
4	1	0	D	0	0
5	1	0	E	0	1
6	1	0	F	1	0
7	1	1	G	0	0
8	1	1	H	0	1
9	1	1	M	1	0

Table A4
Codebook for bit binary grouping

Three bit binary group			Alpha notation	Three bit binary group		
1	2	3		m - 2	m - 1	m
0	1	0	A	0	0	0
0	1	1	B	0	0	0
1	0	1	C	0	0	0
1	1	1	D	0	0	0
1	1	0	E	0	0	0
1	0	0	F	0	0	0
0	1	0	G	0	1	1
0	1	1	H	0	1	1
1	0	1	I	0	1	1
1	1	1	J	0	1	1
1	1	0	K	0	1	1
1	0	0	L	0	1	1
0	1	0	M	1	0	1
0	1	1	N	1	0	1
1	0	1	O	1	0	1
1	1	1	P	1	0	1
1	1	0	Q	1	0	1
1	0	0	R	1	0	1
0	1	0	S	1	1	1
0	1	1	T	1	1	1
1	0	1	U	1	1	1
1	1	1	V	1	1	1
1	1	0	W	1	1	1
1	0	0	X	1	1	1

Three bit binary group			Alpha notation	Three bit binary group		
1	2	3		m - 2	m - 1	m
0	1	0	Y	1	1	0
0	1	1	Z	1	1	0
1	0	1	a	1	1	0
1	1	1	b	1	1	0
1	1	0	c	1	1	0
1	0	0	d	1	1	0
0	1	0	e	1	0	0
0	1	1	f	1	0	0
1	0	1	g	1	0	0
1	1	1	h	1	0	0
1	1	0	i	1	0	0
1	0	0	j	1	0	0
0	1	0	k	0	1	0
0	1	1	l	0	1	0
1	0	1	m	0	1	0
1	1	1	n	0	1	0
1	1	0	o	0	1	0
1	0	0	p	0	1	0
0	1	0	q	0	0	1
0	1	1	r	0	0	1
1	0	1	s	0	0	1
1	1	1	t	0	0	1
1	1	0	u	0	0	1
1	0	0	v	0	0	1

Java source code:

Part of the Java source code which convert the expression into binary equivalent in the case where we have shortening process with two binary pairs. The main class `AN_Notation` contain the method `notation()` which realize Algorithm 1 through four phase:

Step 1 (*Replace*): Converts BP notation into binary equivalent.

```
BP1 = BP.replace("(","1");
BP1 = BP.replace(")","0");
binaryEq = BP1;
```

Step 2 (*Elimination*): Process of elimination of the first and the last bit (`1-0`)

```
BP2 = binaryEq.substring(1,LabelBP.length()-1);
```

Step 3 (*Selection*): Selection of the first and the last two binary pairs of record and we write Alpha notation record based on the table grouping (Table A3)

```
int aLenght = BP2.length();
String aFirst = BP2.substring(0,2);
String aLast = BP2.substring(aLenght-2,aLenght);
String str0="00";
String str1="01";
String str2="10";
String str3="11";
// codebook - table A3 (FROM A TO M)
if (aFirst.equals(str1) && aLast.equals(str0)) Alfa="A";
if (aFirst.equals(str2) && aLast.equals(str0)) Alfa="B";
if (aFirst.equals(str1) && aLast.equals(str2)) Alfa="C";
...
if (aFirst.equals(str3) && aLast.equals(str0)) Alfa="M";
```

Step 4 (*Conversion*): Converts the central part (the rest) into Decimal number.

```
CentralBin = BP2.substring(2,BP2.length()-2);
long numSk = Long.parseLong(CentralBin);
long remSk;
while(numSk > 0){
remSk = numSk % 10;
numSk = numSk / 10;}
int CentDec= Integer.parseInt(CentralBin,2);
CentralDec = Integer.toString(CentDec);
```

An Interaction-based Scenario and Evaluation of Alternative-Fuel Modes of Buses

András Farkas

Faculty of Business and Economics, Óbuda University
Tavaszmező utca 17, H-1084 Budapest, Hungary
e-mail: farkas.andras@kgk.uni-obuda.hu

Abstract. The problem of alternative-fuel modes of buses used for public transportation in urban areas is considered. Classification and characteristics of these vehicles are discussed in detail. A multi-criteria decision analysis (MCDA) method called MAROM is used to rank and evaluate the alternative-fuel modes of buses for an empirical study taken from the literature. Numerical results are compared to those generated from a classical MCDA method called TOPSIS. Exploring causal effects, cross-impacts between the fuel modes are formulated and trend projections are made using a dynamic simulation model. Findings for the expected changes in the characteristics of the alternative-fuel modes in the long-run are then analyzed. Finally, the formal description of the scaling method MAROM is also presented.

Keywords: alternative-fuel modes; multi-attribute decision making; cross-impact analysis

1 Introduction

A characteristic feature of the modern age is the issue of ever growing urbanization. A UN World Urbanization Prospects report projected that 60% (4.9 billion people) of the world's population will live in urban areas by 2030 [37]. As a sharp illustration of this issue, the urban population in India has increased from 62 million (17% of its total population) in 1951 to 285 million (29%) by 2001, and is estimated to grow to around 540 million (37%) by 2021 [23]. Under such circumstances, increasing attention has been given to urban sustainability over the past decades. The sustainable development of cities largely depends upon a sound urban transportation policy that is capable of drastically reducing air and noise pollution in the urban world in order to preserve human health and the environment [15]. Total transport energy use and carbon emission are projected to be approximately 80% higher than current levels by 2030 [12]. Road transport accounts for by 23% of world energy-related CO_2 emissions (e.g., in 2006 it was reported to be 6.3 Giga tons) [19].

This paper focuses on the modern technology and its applicability to mass transit systems as major contributors to sustainable urbanization. The main parameter in defining alternative-fuel solutions is the fuel mode. Worldwide efforts have made for developments and use of alternative-fuels for buses which possess different characteristics than the traditional ones. This issue has attracted immense interest in recent years; see [12], [17], [22], [25] and [34]. Of the various options available for public transportation, efficient bus systems can be effective and affordable.

Tzeng et al. [34] reviewed the most promising developments of alternative-fuel buses suitable for urban areas and compared them to the characteristics of the conventional internal combustion diesel engine bus. They have presented a comprehensive multi-attribute investigation of these alternative-fuel modes with a set of data provided by different groups of Taiwanese experts using the method called TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), which is one of the most popular advanced procedures for evaluating and ranking different alternatives, see [13] and [4]. TOPSIS defines the best option as the one that is closest to the ideal option and farthest away from the negative ideal point.

The author of the present paper has considered another concept in his MCDA methodology called MultiAttRibute Object Measurement (MAROM) [8]. MAROM requires that non-quantifiable and quantifiable attributes be treated in different manners. Raw data, either elicited from experts' judgments or arisen from physical measurements, are preserved for the computations in forms of binary variables, rank numbers and quantitative data depending upon their corresponding scale of measurement, i.e. nominal, ordinal, interval or ratio. This way, they are not subject to any arbitrary transformation onto a higher or a lower order scale. Different units of measurement of the attributes are handled by standardization on those types of scale of measurement where it is necessary.

In this paper we will compare the priority rankings and the performance scores of the alternative-fuel buses resulting from the use of the two methods MAROM and TOPSIS when they are applied to the same data set that has been provided in the seminal paper of Tzeng et al. [34]. Furthermore, we intend to reveal deterministic interactions between the alternative-fuel modes, i.e., between their constituting attributes. Utilizing these cross-impacts, a dynamic simulation model called KSIM [14] will be used to make trend projections for the alternative-fuel modes in order to estimate the changes in their characteristics and in the resulting new ranking over a long-term perspective.

The organization of the article is as follows. In Section 2, a technical description of the alternative-fuel vehicles is presented. In Section 3, the set of criteria and the criteria weights are described. In Section 4, the derivation of the MAROM ranking and the scores of the fuel modes with their comparison to those generated from the TOPSIS are discussed. In Section 5, a cross-impact matrix is constructed, the dynamic simulation model is run and the evaluation of the results is presented. In the Appendix, the formal description of the method MAROM is provided.

2 Characteristics and Trends of Alternative-Fuel Vehicles

Based on the excellent work of Morita [20], first, we present an overview about the characteristics and the main directions of engineering developments of automotive power sources to be expected over the next two decades. We mention here that one of the best sources of information on the development of alternative-fuels can be found in the U.S. Department of Energy's Alternative Fuels Data Center (AFDC) which maintains an "Alternative Fuels Data Base" [1]. Additionally, we refer to the excellent book [3] which treats this subject fully. For grouping public vehicles, Morita [20] suggested to consider four categories of the following types: (i) *internal combustion engine vehicles* (ICEVs), (ii) *electric vehicles* (EVs), (iii) *hybrid electric vehicle*s (HEVs) and (iv) *fuel cell vehicles* (FCVs).

(i) ICEV1 ─ *Vehicles with gasoline engines*

Their heat efficiency exceeds 50%. Main directions of the developments of these conventional vehicles will focus on the drastic reduction of emission pollutants, especially during the cold engine startup and the warm-up periods. For this purpose, they will be installed with three-way catalysts and computerized fuel injection. The adoption of some new technologies, like the use of manifold catalytic converters, the reduction of heat capacity of exhaust systems, the lean-burn operation immediately after engine startup, the use of high precision oxygen sensors and the predictive control of fuel consumption per cylinder have resulted in the production of a series of ultra low emission vehicles whose outputs are only 13% of the current regulation level [20]. Furthermore, a direct-injection engine (e.g. Mitsubishi's GDI engine through a thinner air fuel mixture) has approached the same efficiency as that of the direct-injection diesel engine under low load conditions [16]. The reduction of nitrous oxide (NO_x) occlusion, however, is an issue that seems to be far from resolved.

(i) ICEV2 ─ *Vehicles with diesel engines*

Professional engineers view diesel engines as the most efficient ones of all internal combustion engines (more than 55% heat efficiency). Large efforts have recently been started to significantly reduce their particulate matter (PM) and nitrogen oxide (NO_x) emissions via modifications of the engine mechanism, e.g., by using an inter-cooler turbocharger and cooled exhaust gas recovery (EGR) [20]. New NO_x occluding catalysts have already been developed which are capable of reducing PM and NO_x emissions by more than 80% [33]. This exhaust after-treatment system requires low-sulfur fuel (a cut to 10% of the present level), since sulfur compounds in the fuel poison the catalyst. Such exhaust processing units in diesel engines are very likely to be installed in the near future. Additionally, the installation of a common-rail computerized high-pressure fuel injector will

enhance the controllability of fuel injection and fuel spray [20]. Improvements for diesel engines are likely in both engine systems and fuels, and thus, the negative image of diesel engines as emitters of dirty and harmful materials and chemicals could be lessened.

(i) ICEV3 ─ *Synthetic fuel engine vehicles*

There is a variety of research for synthetic fuels which can be used to power diesel engines. For example, dimethyl ether (DME) can reduce the discharge of soot, since it contains oxygen [11]. Methanol can also be made into ether (MTBE), then blended with gasoline to increase octane. Methanol and ethanol are clear, colorless liquids, i.e., alcohols which are made primarily from natural gas, but also from wood, coal and biomass. These fuels are suited to infrastructures for diesel oil and LPG after a slight modification only [31]. Bio-diesel (mono alkyl esters) is made from natural sources such as vegetable oils and animal fats. Their emission properties are better than those of the conventional diesel engines, since they give a substantial reduction of unburned hydrocarbons (HC), carbon monoxide (CO) and PM. Much of the current interest in bio-diesel production comes from soy-bean producers (e.g., Soy Diesel in the US). Their use as a fuel is intended for bus fleets [18]. The synthetic fuel engines reduce vehicle emissions of pollutants and greenhouse gases significantly.

(i) ICEV4 ─ *Natural gas engine vehicles*

Natural gas (NG) is a mixture of HC, mainly methane (NH_4), produced either from gas wells or together with oil production. It has clean-burning qualities, a wide resource base and commercial availability to users. NG must be stored on a vehicle's board as either compressed natural gas (CNG) or liquefied natural gas (LNG). The on-board fuel storage of cylinders is much stronger than gasoline fuel tanks and is subjected to continuous heat, pressure and fires tests. Natural gas is widely distributed in many countries, e.g., in the US, Canada, Japan and Russia, through extensive pipeline systems. In Japan, by using a ceramic combustion chamber, Isuzu Inst. has achieved a heat efficiency of 39% and they aim to further improve by collecting exhaust energy with a turbine generator to shield heat [21]. NG has numerous benefits, e.g. less pollutants, greenhouse gases, safety and general abundance. These vehicles would work best with mass transit vehicles. Their costs are an average of 15-40% percent less than gasoline and diesel vehicles, including maintenance. The emissions of CO content are 70% and NO_x contents are approximately 88% lower than those of the diesel buses.

(ii) EV ─ *Electric vehicles*

A zero-emission alternative to petroleum is the electric motor driven vehicle, an option currently used in many towns with electric-cable buses. Recent technology, however, uses electricity independently of a fixed electric cable with a fuel cell or battery storage. Its big appeal is having a clean and quiet operating system. Under low-load conditions some metropolises have begun to employ electric buses, but

their future is uncertain, mainly because of the high costs [6]. The key weaknesses of the EVs are the time needed to recharge the batteries, the lack of support infrastructure and the short cruising distance (~200 km). If, however, much bigger batteries are installed, these would add to the vehicles' weight considerably. In general, a major shortage is in this respect that batteries are expensive and, therefore, account for a large proportion of vehicle costs. Rapid-charge batteries with vastly improved energy density may appear only in the distant future. Thereby, battery-electric buses seem to be feasible in low-kilometer circulator routes in the central business district (especially micro EVs) [9].

(iii) HEV1 ─ *Series hybrid electric vehicles*

A series HEV uses the engine driving force after converting it into electricity via a generator. In general, useful properties of HEVs are regeneration of braking energy, engine shutdown instead of idling, and engine driving under high-load conditions (reduction of low-load driving time) [20]. The main advantage of a series HEV is that the engine driving range can be easily optimized compared to other systems, since the engine and the axle are not linked mechanically. However, this system has high energy losses, since the system transmits energy through several modules: the engine, the generator, the battery and the electric motor. The fuel efficiency of a series HEV is better than that of a conventional diesel vehicle under low-speed, low-load conditions, whereas engine efficiency is reduced and average speeds are lower. Therefore, this system is suited to urban driving patterns where stop-and-go at low speeds is common. It will be adopted in city buses in the future.

(iii) HEV2 ─ *Parallel hybrid electric vehicles*

The main attractiveness of a parallel HEV is that it lends itself a better fuel efficiency than a conventional vehicle, regardless of the driving conditions, since it applies the engine's motive force directly during high-load driving. If engine fly-wheels are replaced with relatively low-powered electric motors, such as in Honda's Insight, the mechanism becomes simpler, since the conventional starter and alternator can be replaced with an electric motor [27]. Furthermore, it can utilize the conventional power transmission system between the transmission and the driving wheels, and the use of heavy, costly batteries is then minimized. These advantages and a standardization of manufacturing parts reduce production costs. A good example is Toyota's Crown hybrid which has an extremely simplified hybrid system that uses 42 Voltage for its auxiliary power system (this will be the standard voltage in the European Union).

(iii) HEV3 ─ *Series/parallel hybrid electric vehicles*

The series/parallel HEV is a combination of the above two hybrid systems and has been actively developed and marketed in Japan and in the US. This system performs efficiently regardless of the driving conditions, since it is powered by the electric motor under low-speed, low-load conditions and by the engine under high-

speed, high-load conditions. In many cases, a dedicated hybrid transmission has to be installed (e.g., a CVT that uses a planetary gear system used in Toyota's Prius, and a belt system CVT that is adopted in Nissan's Tino hybrid). The system developed by Hino comprises a very simple design in which one engine and two electric motors are arranged concentrically [38]. The design and control of the transmission are regarded as the key to the success of the series/parallel HEV.

(iv) FCV ─ *Fuel cell (hydrogen) vehicles*

Professionals think of a fuel cell as the ultimate form of public vehicles [20]. Recent interest in hydrogen as a substitute for gasoline and diesel in the transportation market is primarily due to two reasons: (a) hydrogen fuel is essentially limitless, as hydrogen can be gained by electrolyzing water (with the use of renewable energy technologies), and (b) hydrogen fuel is clean-burning, as the oxidation of hydrogen yields only water. Hence, many organizations and governments expect hydrogen to meet a larger share of global energy use in the coming decades [36]. Fuel cells use hydrogen as fuel, but hydrogen is not suitable for on-board storage. Fuel cells generate electricity on board the vehicle in a compact assembly for powering the vehicle from an electrochemical reaction between hydrogen and oxygen under controlled conditions. The only waste in this process is water vapor. The use of indirect hydrogen carriers, such as methanol, gasoline, natural gas and then extracting their hydrogen are currently being studied by researchers [6]. For example, Toyota has introduced a compact methanol reformer which combines an evaporator, an external reformer and a CO eliminator. The internal reformer, which uses direct methanol fuel cells, has a performance output of 6 kW and an efficiency of 40%. Hydrogen's energy density is very low when compared to methanol and especially to gasoline. Therefore, it requires very large, heavy tanks on board which is undesirable for a compact, lightweight vehicle. Researchers agree, however, that using hydrogen as a fuel for mass transit would be advantageous in large buses refueling at a central location. Thus, the on-site NG and a system in which the NG is transported through pipelines to independent refueling stations seem to be the most likely technologies in 15-20 years [36]. In the latter, the NG is reformed to H_2 using an on-site, stationary steam reformer. The H_2 is stored as a compressed gas and dispensed into direct hydrogen FCVs. From a vehicle perspective, this technology configuration is similar to the on-board hydrogen alternative.

Tzeng and his co-authors considered 12 alternative-fuel modes of buses for public transportation in their seminal paper [34]. We will utilize these twelve choices for the types of alternative-fuel vehicles, denoted them as AFV_k, k=1,…,12.

AFV 1: Conventional Diesel Engine—CD

The diesel engine still is one of the major contenders as a power source in the 21st Century. Its main advantages are low purchasing costs, flexibility to the speed of traffic and low sensitivity to road facility. However, it has very high exhaust

emission rates (PM, NO_x, CO, CO_2). This vehicle is introduced in the set of alternatives in order to compare it with the new fuel modes.

AFV 2: Compressed Natural Gas—CNG

Interest for natural gas as an alternative fuel arises from its clean-burning qualities and its wide resource base. Natural gas has numerous benefits in terms of pollutants, comfort, and general abundance. CNG vehicles emit only slight amounts of CO and CO_2, they have high-octane value and they cost less than diesel buses. Meanwhile, natural gas vehicles are saddled with problems in many countries such as supply, distribution and especially risk of explosion.

AFV 3: Liquefied Propane Gas—LPG

There are countries that use this mode of fuel for public transportation. In Japan, Italy and Canada, 7% of transit buses are powered by LPG, and several European countries are planning to employ LPG vehicles, due to pollution considerations.

AFV 4: Fuel Cell (hydrogen) —FC (H)

Research on a fuel cell-hydrogen bus has already been concluded with success. Test results with the experimental vehicle operating on hydrogen fuel indicate that this vehicle has a broad surface in the burning chamber, low burning temperature, and the fuel is easily inflammable. No detrimental substance is produced and only pure water, in the form of vapor, is emitted. A fully loaded fuel tank can last as far as 250 km.

AFV 5: Methanol—MET

The fuel of methanol is related to vehicles with gasoline engines. The combination rate of methanol in the fuel is 85% (so-called M85). The engine can run smoothly with any combination rate of gas with methanol, and methanol will act as an alternative fuel and help to reduce the emission of black smoke and NO_2 as well as pollutants and greenhouse gases. Fuel stations providing methanol are already available in several countries. The thermal energy of methanol is lower than that of gasoline, and the capability of continuous travel by this vehicle is inferior to that of conventional vehicles.

AFV 6: Electric Vehicle - opportunity charging—E-OC

The source of power for the opportunity charging electric vehicle is a combination of a loaded battery and fast opportunity charging during the time the bus is idle. Whenever the bus starts from the depot, its battery will be fully charged. During the 10-20 sec when the bus is stopped, the power reception sensor on the electric bus (installed under the bus) will be lowered to the charging supply plate installed in front of the bus stop to charge the battery. Within 10 sec of a stop, the power supply is done, so that the battery is charged with 0.15 kWh, which is adequate for it to move to the next bus stop.

AFV 7: Direct Electric Charging—E-DEC

The big appeal of electricity is a clean and quiet operating system. This is to be contrasted to its high costs and short cruising distance. The power for this vehicle comes from a loaded battery. Once the battery power is insufficient, the vehicle will have to return to the plant to conduct recharging. The development of a suitable battery is critical for this mode of vehicle.

AFV 8: Electric Bus with Exchangeable Batteries—E-EB

Here, the goals are to accomplish a fast battery charge and achieve a longer cruising distance. The bus is modified to create more on-board battery space, and the number of on-board batteries is adjusted to meet the needs of different routes. The fast exchanging facility has to be ready to conduct a rapid battery exchange.

AFV 9: Hybrid Electric Bus with Gasoline Engine—HE-G

The electric-gasoline vehicle has an electric motor as its major source of power and a small-sized gasoline engine. When electric power fails, the gasoline engine can take over and continue the trip. The kinetic energy rendered during the drive will be turned into electric power to increase the cruising distance of these vehicles.

AFV 10: Hybrid Electric Bus with Diesel Engine—HE-D

The electric–diesel vehicle has an electric motor and small-sized diesel engine as its major source of power. When electric power fails, the diesel engine can take over and continue the trip, while the kinetic energy rendered during the drive will be turned into electric power to increase the cruising distance of these vehicles.

AFV 11: Hybrid Electric Bus with CNG Engine—HE-CNG

The hybrid electric-CNG vehicle has an electric motor and a small-sized CNG engine as its major source of power. When electric power fails, the CNG engine takes over and provides the power, with the kinetic energy produced converted to electric power to permit continuous travel.

AFV 12: Hybrid Electric Bus with LPG Engine—HE-LPG

The hybrid electric-LPG vehicle has an electric motor and a small-sized LPG engine as its major source of power. When electric power fails, the LPG engine takes over and provides the power, with the kinetic energy produced converted to electric power to permit continuous travel.

3 Weighting of Criteria and Evaluation of Alternative-Fuel Modes of Buses

Tzeng et al. [34] used the following 11 single criteria to evaluate the alternative-fuels of buses:

C 1: Energy supply ─ Annual costs of supply, storage and fuel

C 2: Energy efficiency ─ Energy consumption related to fuel heating value

C 3: Air pollution ─ Chemical substance harmful to health

C 4: Noise pollution ─ Noise produced during operation

C 5: Industrial relationship ─ Impact on other locomotive industry branches

C 6: Costs of implementation ─ Costs of production, purchase, implementation

C 7: Costs of maintenance ─ Annual costs of maintenance

C 8: Vehicle capability ─ Cruising distance, gradeability, speed of vehicle, etc.

C 9: Road facility ─ Necessary features of road for the bus, e.g. pavement, slope

C 10: Speed of traffic flow ─ Conformity to traffic flow

C 11: Sense of comfort ─ Traveling comfort and aesthetic appeal

In Table 1, the normalized average weights (relative importance of each criterion) are indicated [34, p.1377]. These weights were determined by several groups of Taiwanese experts using the AHP method [26].

Table 1
Criteria weights and results of the value assessment for the alternative-fuel vehicles [34]

	C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11
Weight	0.0313	0.0938	0.1661	0.0554	0.0629	0.0829	0.0276	0.1239	0.0805	0.1994	0.0761
AFV 1	0.82	0.59	0.18	0.42	0.58	0.36	0.49	0.79	0.81	0.82	0.56
AFV 2	0.77	0.70	0.73	0.55	0.55	0.52	0.53	0.73	0.78	0.66	0.67
AFV 3	0.79	0.70	0.73	0.55	0.55	0.52	0.53	0.73	0.78	0.66	0.67
AFV 4	0.36	0.63	0.86	0.58	0.51	0.59	0.74	0.56	0.63	0.53	0.70
AFV 5	0.40	0.54	0.69	0.58	0.51	0.52	0.68	0.52	0.63	0.60	0.70
AFV 6	0.69	0.76	0.89	0.60	0.72	0.80	0.72	0.54	0.35	0.79	0.73
AFV 7	0.77	0.79	0.89	0.59	0.73	0.80	0.72	0.47	0.44	0.87	0.75
AFV 8	0.77	0.79	0.89	0.59	0.73	0.80	0.72	0.51	0.48	0.87	0.75
AFV 9	0.77	0.63	0.63	0.52	0.66	0.63	0.65	0.67	0.70	0.80	0.74
AFV 10	0.77	0.63	0.51	0.58	0.66	0.63	0.65	0.67	0.70	0.80	0.74
AFV 11	0.77	0.73	0.80	0.48	0.63	0.66	0.65	0.67	0.71	0.62	0.78
AFV 12	0.77	0.73	0.80	0.48	0.63	0.66	0.65	0.67	0.71	0.62	0.78

In Table 1, the averages of the assessed values for the performance of each of the alternative-fuel modes with respect to every criterion are also presented. These values, denoted by u_{ij}, $0 \leq u_{ij} \leq 1$, are taken from [34, p.1378]. They have been derived through conducting a survey by applying a Delphi procedure that was repeated twice [34]. The experts represented manufacturing industries, energy committees, governmental departments, research institutes and academic faculties.

4 Comparison of the Results from TOPSIS and MAROM

The following achievements were first published in [7]. The MAROM procedure (see Appendix) requires that nature of the data for each criterion be adequate to the properties of the type of the scale of measurement to which these data correspond. Therefore, as its first step, each criterion should be assigned to the appropriate scale of measurement. Besides, the number of criteria was extended from 11 to 15, because in the article of Tzeng et al. [34] some additional information is presented which were not directly captured by their analysis. This supplementary data relates to a number of relevant engineering and chemical characteristics of alternative fuels which originated from reliable sources (physical measurements) and are presented in [34, p.1382–1383] with their accompanying units of measurement.

To preserve the uniformity of the two data sets as much as possible (which were used by Tzang et al. [34] and the present author; see Table 1 and Table 2) only minimal changes have been made. This way, criteria C4, C5, C9, C10 and C11 of the original data set have been retained, but they were assigned to ordinal scales so that their original performance values, u_{ij}, see in Table 1, were converted to rank numbers using a nine-grade ordinal scale [1, 1.5, 2, 2.5,… , 5], where an ideally best performance, if there exists any, would receive grade 5.

Utilizing the technical data collected by [34], several new criteria were introduced. As seen in Table 2, these are: ‘Depot’, which can be small or large characterizing the depositary needs of the buses, as a nominal variable [0 or 1], while ‘Cruising distance’, ‘Number of passengers’, ‘Maximum speed’ for urban/suburban services and ‘Recharge time’ are ratio scale variables with specific units of measurements. They constitute the extended form of the “old” criterion ‘Vehicle capability’ whose weight has been uniformly allocated to them. The “old” criterion ‘Energy efficiency’ is a dimensionless variable, since it gives the ratio of the alternative-fuel efficiency/fuel heating value related to that of the diesel bus, and hence, it is reasonable to assign it to an interval scale. For some ratio-scaled criteria, i.e., for the different ‘Costs’, ‘Exhaust emission’ and ‘Recharge time’, obviously, the smaller values represent the better performances.

Table 2
Input data of the alternative-fuel vehicles for MAROM [7]

		AFV1	AFV2	AFV3	AFV4	AFV5	AFV6	AFV7	AFV8	AFV9	AFV10	AFV11	AFV12
Nominal scale	Aggregated weight of nominal scale 0.0666												
	Criterion weight 0.0248												
	Best value on nominal scale for criterion C1 1												
1.Depot		1	0	0	1	1	0	0	0	0	0	0	0
Ordinal scale	Aggregated weight of ordinal scale 0.3333												
	Criteria weights 0.0805 0.0554 0.0629 0.1994 0.0761												
	Best values on ordinal scale for criteria C2-C6 4.0 3.0 3.5 4.5 4.0												
2.Road facility		4.0	4.0	4.0	3.0	3.0	1.5	2.0	2.5	3.5	3.5	3.5	3.5
3.Noise pollution		2.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	2.5	3.0	2.5	2.5
4.Indust. rel.ship		3.0	3.0	3.0	2.5	2.5	3.5	3.5	3.5	3.5	3.5	3.0	3.0
5.Speed of traffic		4.0	3.5	3.5	2.5	3.0	4.0	4.5	4.5	4.0	4.0	3.0	3.0
6.Sense of comfort		3.0	3.5	3.5	3.5	3.5	3.5	4.0	4.0	3.5	3.5	4.0	4.0
Interval scale	Aggregated weight of interval and ratio scales 0.6000												
	Criterion weight 0.0938												
	Best value on interval scale for criterion C7 10.9												
	Worst value on interval scale for criterion C7 0.7												
7.Energy efficiency [dim.less]		1.0	0.8	0.7	1.9	0.8	10.9	5.5	3.2	1.5	1.5	1.5	1.5
Ratio scale	Criteria weights 0.0313 0.1661 0.0248 0.0248 0.0248 0.0248 0.0248 0.0829 0.0276												
	Best values on scale for criteria C8-C15 3875 0.30 500 80 120 10 100000 10410												
	Worst values on scale for criteria C8-C15 46600 30.15 80 40 60 360 624000 30720												
8.Fuel costs [1000 NT$]		14000	11450	15000	46600	14495	3875	4000	8000	7880	7450	7250	7300
9.Exhaust emission ($PM+NO_x+HC+CO_x$) [%]		30.15	19.27	8.2	6.78	12.71	0.30	0.35	0.38	9.97	11.25	10.30	10.55
10.Cruising distance [km]		450	500	400	325	225	100	80	220	250	250	350	350
11.Number of passengers [No]		80	70	70	60	60	50	40	40	50	50	55	55
12..Max speed [km/h]		120	80	90	75	110	60	65	65	70	70	75	75
13.Recharge time [min]		10	200	100	10	10	360	300	300	360	360	360	360
14.Costs of implementation [1000 NT$]		100000	420000	300000	624000	144000	340000	360000	380000	400000	420000	440000	450000
15.Costs of maintenance [1000 NT$]		11400	10410	12500	30720	14700	18495	19600	19200	22200	22400	23500	23800

It is hoped that using this slightly modified data base for the evaluation problem will provide us more robust and reliable results. Table 2 presents the reformulation

of the original data set that meets the requirements of the theory of measurement. In this table, the input data for MAROM, i.e., the characteristic values for the 12 alternative-fuel buses, the 15 single criteria weights and the aggregated weights for the different scales of measurement are indicated.

The results of the multi-criteria evaluation of the 12 alternative-fuel buses are shown in Table 3. Here, both the ranks and the evaluation indices called relative standings (scores) yielded by using TOPSIS (basic and compromise solutions) and MAROM (for the individual and the aggregate weighting cases) are indicated. The scores appear on [0 –1] interval scales.

Table 3
Comparison of the rankings and the evaluation scores for TOPSIS [34] and MAROM [7]

	TOPSIS				MAROM			
	Rank Basic	Score	Rank Compr.	Score	Rank Indiv.	Score	Rank Aggreg.	Score
Electric bus with exchangeable batteries	1	0.945	1	0.975	5	0.514	4	0.675
Electric bus with opportunity charging	2	0.933	3	0.964	7	0.498	3	0.677
Electric bus with direct charging	3	0.931	2	0.967	4	0.514	2	0.681
Hybrid electric with gasoline engine	4	0.749	9	0.756	9	0.482	7	0.630
Hybrid electric with CNG engine	5	0.700	4	0.889	11	0.449	11	0.599
Hybrid electric with LPG engine	6	0.700	5	0.889	12	0.448	12	0.599
Hybrid electric with diesel engine	7	0.700	11	0.488	8	0.484	8	0.629
Fuel cell (hydrogen)	8	0.563	6	0.865	3	0.733	10	0.601
Methanol	9	0.527	10	0.698	1	0.791	1	0.691
Compressed natural gas engine (CNG)	10	0.399	7	0.830	10	0.467	9	0.611
Liquidate propane gas engine (LPG)	11	0.345	8	0.830	6	0.499	5	0.670
Conventional diesel engine bus	12	0.301	12	0.097	2	0.785	6	0.650

As it does not come as a surprise, the two methods have produced rather different rankings and scores. Comparisons of the findings, however, should be made very carefully. As a remarkable outcome, observe the big differences in the ranks of the conventional diesel engine bus. The last position of the diesel engine in the TOPSIS rankings seems to be rather strange regarding the fact that Tzeng et al.'s investigations refer to the year 2005. It is also striking that there are significant differences in the priority scores of the alternative-fuel modes produced by the two methods. We intend not to go into detailed technical explanations, only to mention that we believe that the MAROM ranking reflects better the situation existing at that time than that of TOPSIS. The relative high positions of the conventional diesel engine bus in the MAROM rankings as opposed to those obtained for the

alternative-fuel modes follows mainly from the tardiness of the required engineering developments and the limited bus manufacturing capabilities as well as the weak achievements of the civil initiatives concerning environmental protection. However, there is no doubt as urban mass transit technology gets stronger and improves, more buses will be powered by alternative means in the search for more efficient energy use, cleaner air, quieter operation, more safety and more traveling convenience, especially, if they could efficiently serve in suburban areas as well.

5 An Alternative-Fuel Mode Scenario for Buses and Its Evaluation

Hereafter we will consider each alternative-fuel mode as being a quantity Q (a compound construct, as each AFV constitutes 15 variables). These attributes are listed in Table 2. Furthermore, the set of the alternative-fuel buses are regarded an interrelated complex system as it is apparent that in the course of their evolution they interact with one another. At this point, a question of vital importance can be raised. Namely, how the characteristics of these alternative-fuel modes will change and what will their spread in public transportation means look like over the successive two decades. To make technology assessments and study the dynamic behavior of this system, we now attempt to employ a dynamic model based on simulation [14].

Kane procedures for modeling such systems require that we specify a set of quantities Q (AFVs in our case); a set of binary interactions C, between any two pairs (q_i,q_j), $i,j=1,\ldots,n$, including possible self-interactions, defined on $Q\times Q$; and a set of initial values for each of the quantities q_i, denoted as q_{i0} (individual scores from MAROM as given in the third column of Table 3). This model conforms well to our problem, since all variables q_i are bounded $0 < q_i(t) < 1$ for all $i=1,\ldots,n$, and for all $t>0$, and, thus, no rescaling is needed. The projected trends of the variables (AFVs) are of sigmoidal type as the solution of the following differential equation (for small Δt time increments, i.e., for one iteration in the simulation):

$$\frac{dq_i}{dt} = -\sum_{j=1}^{n} c_{ij} q_i q_j \ln q_i, \qquad i = 1,..., n,$$

where c_{ij} is a binary interaction coefficient of q_j upon q_i. From this equation, it becomes clear that q_i accumulates the effect of q_j, since it is easy to see that [2]:

$$q_i(t) = q_{i0} + \int_{j=1}^{n} f[q_j(\tau)]d\tau, \qquad i = 1,..., n.$$

Observe that the structure of the model does not imply that the interaction coefficients were constants. Their behavior is just the contrary, since they are varying with time; see [10] for more details. Properties of the KSIM simulation model are as follows [14]:

(i) System variables are bounded. In an appropriate set of units these can always be set to one and zero.
(ii) A variable increases or decreases according to whether the net impact of the other variables is positive or negative.
(iii) A variable's response to a particular impact decreases to zero as that variable approaches its upper or lower bound. Bounded growth and decay processes exhibit a sigmoid type character.
(iv) All other things being equal, a variable will produce greater impact on the system as it grows or it declines larger.
(v) Complex systems are described by a network of binary interactions.

In order to gain insight into the conjectures of pairwise causal relationships of deterministic type between the AFVs, the actual interaction coefficients were revealed following extensive research of related literature and juries of executive opinion. If the c_{ij} coefficients are generated through some kind of Delphi procedure, the respondents, in justifying a nonzero assignment to a particular c_{ij} may provide an intuitive argument for, and thus a possible explanation of, the claimed casual relationship. While the c_{ij}'s individually are not exploratory in nature, the matrix called cross-impact matrix, $C=[c_{ij}]$, as a whole represents a coherent pattern of causality assertions. For our case, it has the form:

$$C = \begin{array}{c|cccccccccc} & \textbf{CD} & \textbf{CNG} & \textbf{LPG} & \textbf{FC} & \textbf{MET} & \textbf{ELO} & \textbf{DEL} & \textbf{HGD} & \textbf{HCL} & \textbf{OW} \\ & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \textbf{CD} & + & - & 0 & - - & 0 & - & - & + + & 0 & - - \\ \textbf{CNG} & - & + & 0 & - & 0 & 0 & 0 & 0 & + & + + \\ \textbf{LPG} & - & 0 & + & - & 0 & 0 & 0 & 0 & + & + \\ \textbf{FC} & + & 0 & 0 & + + + & 0 & 0 & 0 & 0 & - & + + + + + \\ \textbf{MET} & 0 & - & - & - & 0 & 0 & 0 & 0 & 0 & 0 \\ \textbf{ELO} & - - & - - & - - & - & - & 0 & 0 & + & + & - \\ \textbf{DEL} & - - & - - & - - & - & - & + + & + & 0 & 0 & - \\ \textbf{HGD} & + + & - & - & - - & 0 & + & + & + & 0 & + \\ \textbf{HCL} & - & + & + & - - & 0 & + & + & - & + + & + \end{array}.$$

To construct matrix C, interviews with several experts' groups (formed from the researchers of different faculties of transportation engineering at the Technical University of Budapest) were conducted and repeated three times to achieve a compromise decision to confirm structure. To simplify our system we have merged similar AFVs having the same individual scores: Direct electric charges ─ Electric bus with exchangeable batteries (DEL); Hybrid electric bus with gasoline

– diesel engines (HGD) and Hybrid electric bus with CNG – LPG engines (HCL). In matrix C, at each cell, the action of (a positive change in) the column heading upon the row heading is entered. A nonzero diagonal complies with the idea that technology tends to foster its own growth usually (self-interaction). Notice that the entries of this matrix are combinations of pluses and minuses, rather than numerals to accentuate the subjective nature of the judgmental procedure. For the strength and direction of the interactions they indicate a weak, or a slight, or a moderate, or a firm or a strong positive and/or a negative impact, respectively. Observe that the cross impacts are not necessarily symmetric. As a unique feature of this model, an extra column for expected actions from the outside world (OW) is also added for external interventions, new international standards and regulations, users' acceptance, civil campaigns, etc. The entries in matrix *C* can be subjected to debate, and then, one would argue for different choices. Therefore, we stress that our cross-impact matrix *C* defines only one possible option for a well-established system of interactions of the AFVs.

Figure 1 exhibits the projected trends of interactions as given by matrix *C* over a two decade time horizon after completing 50 simulation runs. Tendency and the changes in the alternative-fuel modes could be analyzed from this scenario by keeping in mind that any change in the behaviour of an impacted AFV is a result in the common effect of its self-development and the changes in its constituting variables caused by the total impact of the changes in the impacting AFVs, as well as different external factors, i.e., new international regulations, users' concerns, etc. As is seen in Figure 1, by 2030, the Fuel cell (hydrogen) bus will take over the "lead" before the Hybrid electric buses operating with gasoline/diesel engines and the Hybrids with CNG/LPG engines. The performance score of the diesel bus will decline significantly. Similar decays can be observed for the CNG and LPG buses.

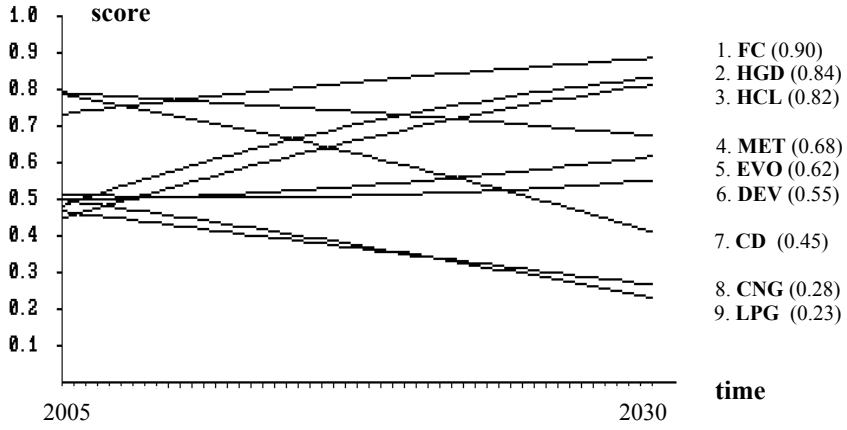


Figure 1
Projected trends and scores from the interactions as given by matrix *C*

The projected trends of this model in Figure 1 reflect the international directions in terms of both technological developments and environmental protection fairly well. Japanese manufacturers are seriously investing in hybrids, what they see as a promising market segment. US manufacturers are starting to use hybrids to disguise the environmental impacts of vehicles that consumers want. European manufacturers respond to the market's need for high performance and less polluting vehicles by investing in diesel technologies and tend to ignore or dismiss hybrid technology as an overly complicated half-solution that introduces excess weight and hampers performance. Instead, they intend to favour fuel cell research as the way forward. In their view, hybrid buses are only a medium-term interim solution filling the gap until a more efficient technology, ideally fuel cell buses mature and become available [32]. It should be mentioned that the struggle to reduce vehicle emission has strongly speeded up recently. As an illustration for this, the bus emission standards for NO_x and PM in the US and in the EU have become much more rigorous between 2000 and 2010, i.e., the NO_x emission in [g/kWh] should be reduced from 5.8 down to 0.16 and from 5.0 down to 2.0, respectively and the PM emission from 0.075 down to 0.0075 and from 0.1 to 0.02, respectively [35]. Technical limitations of electric and hybrid vehicles are mainly related to capacity, durability and price conditions of the batteries. In summary, our projections for AFVs are fairly close to those scenarios reported by the European Commission [24], with the small exception that the CNG vehicles' performance and thus the popularity of their use in public transport seem to be a little bit under-estimated by the year 2030.

Conclusions

In this paper, current status and future developments of alternative-fuel modes of buses have been reviewed. Two MCDA scaling methods, TOPSIS and MAROM, have been compared by applying them to the same empirical example, the qualification of alternative-fuel buses operating in urban areas. Although the results produced by the two methods were quite different, both approaches have shown that the use of alternative-fuel modes to improve human health and the environment offers huge opportunities for ensuring sustainable urban development. An interaction-based simulation model was applied to the system of alternative-fuel modes of buses to project the expected changes in the main characteristics of these public transportation means. We conclude that the technological developments of hybrids look positive, but their market potential is quite questionable. However, hybrids can be an important supplementary technology. Policy makers are seeking for long-term technological solution to decrease fuel consumption, dependence on petroleum, carbon dioxide, carbon monoxide and other exhaust fumes emission while seriously investigating the options for the transition to hydrogen-based technologies in the long-term.

Appendix

The formal description of the method MAROM is presented below:

Consider the following data matrix:

$$A = [a_{ik}], \qquad i = 1,2..,m; \ k = 1,...,n, \tag{1}$$

involving n options (alternatives). The n columns give for every option the values of m variables (row vectors) denoting various characteristics (attributes, criteria) of these alternatives. In (1), a value (crisp number) is assigned to each entry a_{ik} which is either elicited from respondents' judgments or arisen from physical measurements. Thereby, the nature of a particular data may be of a subjective type (qualitative) and/or an objective type (quantitative). A column vector a_k of matrix A represents a composite vector $a_k=(a_k^{(N)}, a_k^{(O)}, a_k^{(I)}, a_k^{(R)})$ which is partitioned into four blocks. Thus, A consists of variables of mixed type, where the superscript N refers to nominal (usually binary), O to ordinal, I to interval and R to ratio variables. Of course, in a concrete real-world case, variables of any type may be missing.

An additional column vector, denoted by b, called a reference vector, is to be constructed which represents an ideal (hypothetical) option, entries of which are composed of the "best" values of the set of alternatives with respect to each attribute. It has the same element-wise structure as that of vector a_k. Applied numerical scales are: nominal: [0,1]; ordinal: [1,…,5]; (linear) interval and/or ratio scales: [actual data emerging from measurements].

Because the ratio scale (and sometimes interval) variables have usually different units of measurements the row vectors $a_i^{(R)\mathrm{T}}$ (and $a_i^{(I)\mathrm{T}}$) are standardized so that their means are equal to 0 and their standard deviations are equal to 1. E.g., for the ratio variables, these standard deviations can be obtained as

$$s_i^{(R)} = \sqrt{\frac{1}{n-1}\left[\sum_{i=1^{(R)}}^{k^{(R)}} a_{ik}^{(R)2} - \frac{1}{n}\left(\sum_{i=1^{(R)}}^{m^{(R)}} a_{ik}^{(R)}\right)^2\right]}, \qquad i = 1^{(R)},...,m^{(R)}; \quad k = 1,...,n. \tag{2}$$

With (2), the standardized elements are

$$a_{ik}^{'(R)} = \frac{1}{s_i^{(R)}}\left(a_{ik}^{(R)} - \overline{a}_{i.}^{(R)}\right), \qquad i = 1,...,m; \quad k = 1,...,n. \tag{3}$$

A representative group of respondents (experts, customers, users, etc.) is then formed. Every committee member should evaluate each alternative by supplying his judgments on each qualitative variable with respect to the nominal and ordinal scaled criteria. It is recommended that the number of voters l, l =1,…,q, to be at least 10 persons.

The multi-attribute decision making model for preference measuring is as follows

$$\overline{d}_k^{\,l} = \sum_{i=1}^{m} w_i^l d_{ki}^{\,l} + \varepsilon_k^{\,l}, \qquad k = 1,...,n; \quad l = 1,...,q, \tag{4}$$

where $\bar{d}_k^l$ is the overall distance of alternative k from the "ideal" alternative for the lth voter; w^l_i is the weight of attribute i; d_{ki} is the distance of the kth alternative (object) from the reference point on attribute i; ε_k is the value of an error random variable which may include model misspecification, measurement errors and respondents' uncertainties. To determine the weights of the attributes, w^l_i, i=1,…,m, the analytic hierarchy process (AHP) method is proposed [26]. These weights are then usually normalized, so that $\sum^n_{k=1} w^l_i = 1$.

The distance measure d_{ki} in Eq. (4) takes on different functional forms for alternative k:

(a) For the nominal vectors, $d_{ki}^{(N)}(a_{ki}^{(N)},b^{(N)})$, denoting them simply as $x,y \in N$, the distance measure is the Tanimoto (also called Jaccard) coefficient [29]:

$$d_{ki}^{(N)}(x,y) = 1 - \frac{\alpha}{\alpha+\beta+\gamma} = \frac{\beta+\gamma}{\alpha+\beta+\gamma},$$

where

$$\alpha = \sum_i \min(x_i, y_i), \quad \beta = \sum_i x_i - \alpha, \quad \gamma = \sum_i y_i - \alpha, \quad i \in N.$$

(b) For the ordinal vectors $d_{ki}^{(O)}(a_{ki}^{(O)},b^{(O)})$, denoting them simply as $x,y \in O$, the distance measure is the Soergel number [28]:

$$d_{ki}^{(O)}(x,y) = \frac{\sum_i x_i + \sum_i y_i - 2\sum_i \min(x_i, y_i)}{\sum_i x_i + \sum_i y_i - \sum_i \min(x_i, y_i)}, \quad i \in O.$$

(c) For the interval vectors and the ratio vectors, $d_{ki}^{(I,R)}(a_{ki}^{(I,R)},b^{(I,R)})$, denoting them as either $x,y \in I$, or $x,y \in R$ and introducing the L_2 norm of a vector x,

$$\|x\|_2 = \sqrt{\sum_i x_i^2} = \sqrt{x^{\mathrm{T}} x}, \quad i \in I, \quad \text{or} \quad i \in R,$$

the distance measure is the well-known Euclidean-metric:

$$d_{ki}^{(I,O)}(x,y) = \|x - y\|_2 = \sqrt{(x-y)^{\mathrm{T}}(x-y)}.$$

Since the metric properties hold for the above distance functions (a), (b) and (c) used in model (4) (see the proofs in [30]), therefore, the additive type composite vector $\bar{d}_k^l$ is also metric. Furthermore, it is unique and for each of the partial vectors of the distances: $0 \le d_{ki}(a_k^{(\cdot)\mathrm{T}}, b^{\mathrm{T}}) \le 1$. The distance between any two composite vectors is proportional to the degree of intensity. The proportionality unit is taken to be 1.

Once the pairwise distances between each composite vector and the reference vector have been determined, then, with the (column) vector of the relative standings (scores), denoted as $s=(s_k)$, k=1,2,...,n, the overall priority ranking of the alternatives yields as the order given by the components: $s_k = 1 - d_k$. To establish

either a [0–1] or a [1–100] interval scale for the overall priority ranking, a simple normalization procedure should be performed. To aggregate the individual rankings of the decision makers into a compromise ranking, the minimum variance method [5] is employed.

References

[1] Alternative Fuels Data Base. U.S. Department of Energy's Alternative Fuels Data Center (AFDC). http://www.afdc.doe.gov/amfa

[2] Burns,J.R. and Marcy,W.M., "Causality: Its characterization in system dynamics and KSIM models of socioeconomic systems." *Technological Forecasting and Social Change*. **14**, (1979), 387-398

[3] Bus System for the Future. Achieving Sustainable Transport Worldwide. International Energy Agency. OECD/IEA, Paris, 2002

[4] Chen,S.J. and Hwang,C.L., 'Fuzzy Multiple Attribute Decision Making: Methods and Applications. Springer Verlag. Berlin, 1992

[5] Cook,W.D. and Seiford,L.M., "On the Borda-Kendall consensus method for priority ranking problems", *Management Science.* **28**, (1982), 621–63

[6] Dzurik,A., Leszczynska,D. and Brenner,A., 'Mass Transit and Sustainable Urban Environments. Urbanicity. 2005. http://www.urbanicity.org/Site/Articles/Dzurik.aspx

[7] Farkas,A.,"A comparison of MCDA techniques TOPSIS and MAROM in evaluating bus alternative-fuel modes", Proceedings of the 11th International Conference on Management, Enterprise and Benchmarking, MEB'13, Budapest, Hungary, May 31-June 1. Óbuda University, (2013), 181-194

[8] Farkas,A., "Priority ranking methods: A survey and an extension", in: Business Research and Management Challenges. (ed. Peter.S), Intl. Mgmt. Center, Budapest, (1994), 74-94

[9] Fowler,T.M., Euritt,M.A. and Walton,C.M., 'Electric Bus Operations: A Feasibility Study. University of Texas Austin, May, 1995, p. 74

[10] Gur,Y., "An extension of structural modeling." *Technological Forecasting and Social Change*. **14**, (1979), 399-408

[11] Hikino,K., "Research and development of a hybrid bus with a DME fueled engine." *JARI Research Journal*. **22**, (2000), 35-38

[12] Hu,A.H., Chen,S.H., Fan,C.H., Hsu,C.W. and Tzeng,G-S., "Evaluation framework for alternative fuel vehicles: Sustainable development perspective." *Energy Policy*. (under review)

[13] Hwang,C.L. and Yoon,K., 'Multiple Attribute Decision Making─Methods and Applications. Springer. New York, 1981

[14] Kane,J., "A primer for a new cross-impact language─KSIM." *Technological Forecasting and Social Change*. **4**, (1972), 129-142

[15] Kazimi,C., "Evaluating the environmental impact of alternative-fuel vehicles". *Journal of Environmental Economics and Management*. **33**, (1997), 163-165

[16] Kimura,S., "Characteristics and improving ways of thermal efficiency on DI diesel engines." *Journal of Society of Automotive Engineers of Japan*. **54**, (2000), 56-61

[17] Lin,C.W., Chen,S.H. and Tzeng,G.H., "Constructing a cognition map of alternative fuel vehicles using the DEMATEL method." *Journal of Multicriteria Decision Analysi*s. **16**, (2009), 5-19

[18] Lynch,T.A., Eliason, L. and Dzurik,A., 'Energy and Environmental Performance of Existing Public Transportation Technologies.' USDOT Research Report. DTRS93-G-0019 – NUT14-FSU4. Florida State University. College of Engineering, Tallahassee, FL, 2004

[19] Metz,B., Davidson,O.R., Bosch,P.R. and Dave,R.,M., "Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. Cambridge, UK and New York, NY, USA, 2007

[20] Morita,K., "Automotive power source in 21st century". *Journal of the Society of Automotive Engineers of Japan*. **24**, (2003), 3-7

[21] Nakashima,K., Kawamura,H. and Matsuoka,H., "Development of a hybrid vehicle with a hear insulated natural gas engine". *JARI Research Journal.* **22**, (2000), 23-36

[22] Offer,G.J., Howey,D., Contestabilec,M., Clague,R. and Brandon,N.P., "Comparative analysis of battery electric, hydrogen fuel cell and hybrid vehicles in a future sustainable road transport system." *Energy Policy*. **38**, (2010), 24-29

[23] Padam,S. and Singh,S.K., "Urbanization and Urban Transport in India: The Sketch for a Policy." Cent.Inst. of Road Transport. Pune, India. World Bank Development Studies, 2002

[24] Proposal for a directive on the deployment of alternative fuels infrastructure. European Commission Staff Working Document. SWD 6. Part I., Brussels, 2013

[25] Romm,J., "The car and fuel of the future." *Energy Policy*. **34**, (2006), 2609-2614

[26] Saaty,T.L., "A scaling method for priorities in hierarchical structures". *Journal of Mathematical Psychology.* **15**, (1977), 234-281

[27] Shimada,A., Ogawa,H., Nakajima,M. and Shimada,H., "Development of the ultra-thin DC brushless motor for hybrid car ─ INSIGHT. *Honda R&D Technical Review*. **12**, (2000), 15-20

[28] Soergel,D., "Mathematical analysis of documentation systems". *Information Storage Retrieval.* **3**, (1967), 129-173

[29] Sokal,R.R. and Sneath,P.H.A., 'Principles of Numerical Taxonomy'. Freeman and Co, San Francisco, 1963

[30] Späth,H., 'Cluster Analysis Algorithms.' Wiley, New York, 1985

[31] Sperling,D., 'Future Drive ─ Electric Vehicles and Sustainable Transportation.' Island Press. Washington DC. 1995

[32] Status and prospects of hybrid technology and the regeneration of energy in road vehicles. Technical Report, EUR 21743 EN. European Commission. Joint research Centre (DG JRC) Institute for Prospective Technological Studies of European Communities. 2005

[33] Tanaka,T., "After-treatment systems and fuel properties for controlling engine emissions." IWPS 2000. (2000), 58-67

[34] Tzeng,G-H, Lin,C-W. and Opricovic,S., "Multi-criteria analysis of alternative-fuel buses for public transportation". *Energy Policy.* **33**, (2005), 1373-1383

[35] Vehicle Emission Reductions. European Conference of Ministers of Transport, OECD /ECMT, Paris, 2001

[36] Winebrake,J.J. and Creswick,B.P., "The future of hydrogen fueling systems for transportation. An application of perspective-based scenario analysis using the analytic hierarchy process." *Technological Forecasting & Social Change.* **70**, (2003), 359-384

[37] World Urbanization Prospects, the 2011 Revision. United Nations, Dept. of Economic and Social Affairs. Population Division. Working Paper. http://www.un.org/esa/population/publications/WUP2011/2011wup.htm

[38] Yokota,H. and Kakegawa,T., "Low-emission diesel-electric hybrid truck by means of new combustion concept and after-treatment technology". *Journal of the Society of Automotive Engineers of Japan*. **56**, (2002), 84-89

General Parametric Model of The Body of The Total Hip Endoprosthesis

Slobodan Tabaković, Milan Zeljković, Aleksandar Živković

Faculty of Technical Sciences, University of Novi Sad
Trg D. Obradovića 6, 21000 Novi Sad, Serbia
e-mail: tabak@uns.ac.rs, milanz@uns.ac.rs, acoz@uns.ac.rs

Abstract: The endoprostheses of the hip joint are the most commonly used group of implants in orthopedic surgery. The success of their application in the process of replacing natural hip joint with artificial depends, besides the patient related and surgical factors, which include the selection of the type of endoprosthesis and the extent of its adaptability to the patient. Because of that, endoprostheses adapted to the patient's measures, so-called custom made endprostheses, are used. The main challenge in their design is large number of influential factors on the shape and dimensions of the endoprosthesis, the long period of development, and high prices as a consequence of these factors.
One of the directions of endoprosthesis design process improvement is the introduction of generalized computational models that enable the adaptation of the endoprosthesis to a specific patient's femur, taking into consideration all influential factors of the disease.
This paper describes a general parametric model of the body of the hip endoprosthesis, which was developed by combining two methods of parametric modeling with the aim to implement all existing contributing factors and, also, facilitate introduction of the new ones. The main advantages of this model of endoprosthesis are: flexibility and simplicity which makes this model easy for application in a variety of CAD software systems.

Keywords: parametric model; endoprosthesis; Bezier curves; Computer Aided Design

1 Introduction

Biomechanical researches have shown that some of the bones and joints of the human skeletal system are exposed to considerable strain. This is the result of daily activities in which the average person makes about 10,000 steps [17]. This may, with the influence of various illnesses or injuries, be a source of various degenerative changes in the hip joint elements (Fig. 1), which establish a connection between the upper part of the skeletal system and the femur.

When damage from a disease becomes severe, it is necessary to perform partial or complete (total) hip joint replacement.

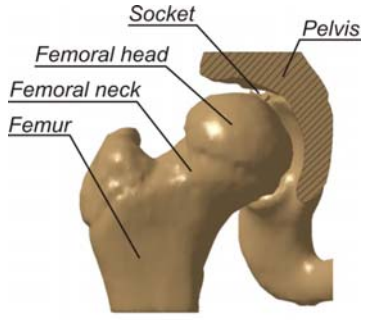


Figure 1
Elements of the hip

The success of the surgery to replace a natural hip joint with the artificial one depends, in addition to the patient related and surgical factors, on so called endoprosthesis factors, which include selection of the type of endoprosthesis and the extent of its adjustment to the patient.

From the design point of view, the primary goal of the endoprosthesis development is to reduce its impact on the success of the surgery. In the last 70 years, since 1938 when the total hip replacement surgery development started [2], a whole range of types of prostheses was developed, which are divided according to their purpose (primary, revision and reconstructive) [7], dimensions (usually there are about 10 models per type) [10], and the way in which the femur is fixed (cement, cementless) [7]. The choice of such standardized endoprosthesis for a patient is made based on the type and complexity of the disease, age of the patient and the basic geometrical parameters of the femur. The main disadvantage of this method of endoprosthesis design is the small number of geometrical factors considered, while some important factors such as the degree of disease and patient ethnicity in general are not considered [10, 16].

According to the current research [9, 11], further development of total hip endoprosthesis can be achieved through the development of implants adjusted to the individual characteristics of the patient. In this way, it is possible to include a number of influential factors and achieve the optimum shape and size of the endoprosthesis. The basic prerequisite for this approach is the improvement of methodologies to determine the characteristic geometrical size of the femur [5, 3] and improvement of the endoprosthesis design process itself.

The endoprosthesis design phase, in which all influential factors are implemented in the computer model, can be realized by parameterization of endoprosthesis geometry elements and by defining the relationship between influencing factors and geometrical parameters. The particular geometry defined in this way is processed by using parametric modeling CAD software systems. For the purposes of complex modeling, two methods of defining the parametric model are used.

These are modeling by using of:

- discrete and
- functionally dependent geometrical parameters.

Modeling by using discrete geometrical parameters is the process of modeling of the product geometry, which in the first phase includes the previous system analysis and decomposition of the model geometry of the endoprosthesis to simple geometrical shapes. In the second stage, the individual simple geometrical shapes and their mutual constraints are defined [8]. Integral parts of this model type are discrete parameters given in general numbers that describe geometrical shapes [7].

Previous studies related to the hip endoprostheses design commonly used modeling employing discrete geometrical parameters principle. The main reasons for this were ease of use and the relatively small number of significant factors (mostly geometrical). On the other hand, the biggest disadvantage of parametric models (based on the parameterization process) is the rigid structure that prevents modifying the dependencies between parameters.

The second method, which uses functionally dependent modeling of geometrical parameters, is based on the functional description of mathematical model parameters. It uses mathematical laws that describe: interrelationships between geometrical parameters [14], or spatial areas. For the definition of complex spatial models polynomial and rational Bezier curves[1] and similar functions are commonly used. This mathematical method of defining the parameters in the model is of a generalized form. This means that the actual endoprosthesis model describes the coordinates of the characteristic points and additional coefficients that have functional dependencies associated with the influential factors in the model. The relationship between the characteristic points and the parameters that define the model is achieved by analytical expressions that may not be contained in the model. Application of functionally dependent geometrical parameters has several advantages over discrete modeling based on geometrical parameters. This primarily includes a significantly more flexible form of the description of geometry, because the model defined is of general nature and can be used for designing different types of endoprostheses. In addition, such models have less geometrical elements that describe the endoprosthesis, which is important from the point of view of speed and accuracy of modeling.

This paper describes the original general parametric model of endoprosthesis body developed by using the functional parametric modeling method which is the result of the previous analysis of geometry of the femur, surgical techniques and a number of existing structural solutions for bodies of the total hip endoprosthesis. Besides the functional dependencies that define geometrical elements of the endoprosthesis body, the specific constraints were additionally introduced in order to simplify the model.

[1] Bezier surfaces were developed by Pier Bezier for the Renault car factory in early sixties [12].

2 Materials and Methods

Design and manufacturing of the total hip endoprosthesis depends on several groups of influencing factors. The most important are: the type and degree of illness, the elements of the femur geometry and the characteristics of surgical techniques used in replacement of a natural with an artificial hip joint. Among them, the largest group consists of geometrical quantities describing the geometry of the femur and its location in the skeletal system, which is important to assure that the patient can walk correctly and smoothly after surgery.

From the design point of view, the endoprosthesis body (femoral stem), as the most complex element of the hip joint endoprosthesis, is a geometrical entity that globally has the same elements. This is the main motivation for the introduction of a unique geometrical form that describes the various forms of body endoprosthesis. This form, a general parametric model of the endoprosthesis body, is a series of interconnected mathematical laws that describe the geometry of the endoprosthesis through the implementation of all influencing factors.

2.1 The Geometry of the Femur

The femur is the longest and, according to its mechanical properties, the strongest bone in human body [10]. Its geometrical properties can be viewed through the position and shape of the exterior and / or interior surfaces. The most important elements of the upper half of the femur (important for the design of endoprosthesis) are (see Fig. 2) [16]: Head (1), Neck (2), Trochanter greater (3), Trochanter lesser (4), the body of the femur (corpus femoris) (5) and femoral (or medullary) channel (6)

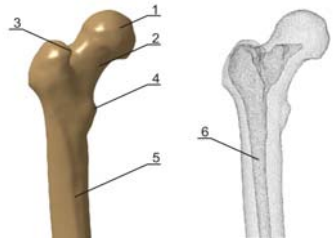


Figure 2
Femur geometry elements

In the anthropometry of the lower extremities of the human body, a group of geometrical values that are influential factors in the design of endoprosthesis [9] are used to describe the position and size of individual elements of the femur, as well as its position in the skeleton. Due to the position of the femur in the skeletal system, these values are described in characteristic anatomical planes: axial (Ap), coronal (Cp) and sagittal (Sp) (Fig. 3).

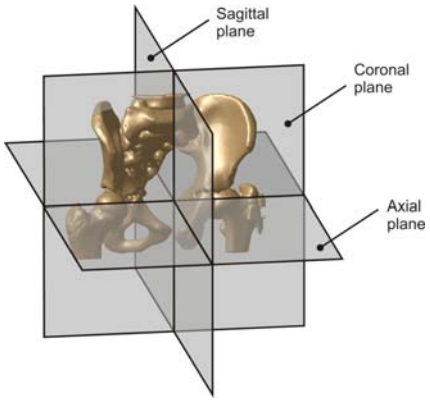


Figure 3
Basic anatomical planes

The characteristic values in the coronal plane are (Fig. 4):

- The angle of inclination of the femur body
- The length of the femoral neck (P)
- Position of the center of femoral head
- The distance between the center of the femoral head and the femur axis (A_1)
- The diameter of the femur head (B)
- Distance between the femoral head from the lesser trochanter (C)

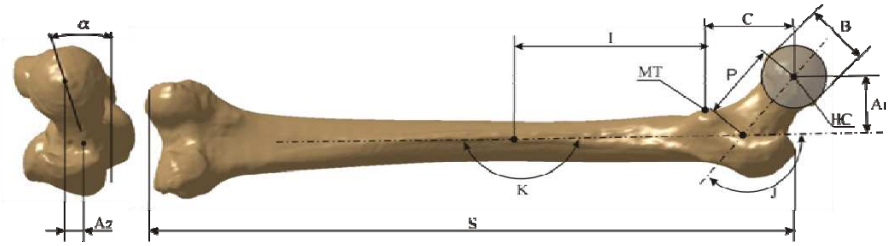


Figure 4
Influential factors in the coronal plane

Another group of influential factors, originating from the geometry of the femur, is defined in the axial plane. It includes:

- Anteversion, the angle between the axis of the femoral neck and coronal planes – α (Figure 4)
- Position of the center of the femoral head (A_2)

Among the geometrical values which determine the shape and dimensions of the internal geometry of the femur i.e. the medullary channel, the most commonly used are as follows (Fig. 5):

- The position of the medullary channel isthmus (G)
- The expansion coefficient of the medullary channel (Channel flare index - CFI), which is determined by the quotient of the channel width in the planes that are positioned in the coronal plane, 20mm below (F) and above (D) of the lesser trochanter, equation (1) [12].

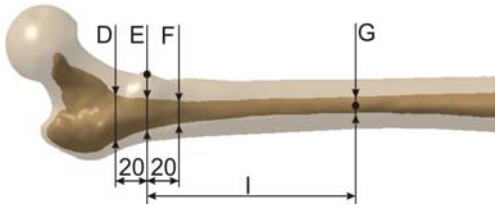


Figure 5
The influential factors of the internal geometry of the femur

$$CFI = D/F \tag{1}$$

The above geometrical quantities determine the position that the endoprosthesis should have in the femur, the global dimensions of endoprosthesis individual body segments and their mutual relationship. As such, they make the basis for a detailed sizing of certain segments of the endoprosthesis.

In addition to the basic geometrical factors, usage of the diagnostic devices based on tomography methods of recording allows introduction of additional factors. They are consequence of the medullary channel shape and influence the process of surgical procedure. These are:

- Dimensions of the cross sections in several characteristic planes (or explicit mathematical expression) (d_{MCAP}, d_{MCLM} on Fig. 6), [12]
- Cement layer thickness (S on Fig. 6).

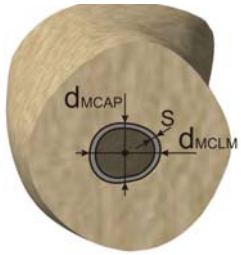


Figure 6
The cross-section of the femur

2.2 Appearance and Endoprosthesis Body Elements

As a result of endoprosthesis body shape standardization, for many years, there are categories of hip joint endoprostheses that are grouped according to the type of disease and the fixation mode of the endoprosthesis into the femur. This approach is verified by the ISO 7206-1/2008 [7], so that endoprostheses are categorized into following types:

- *Primary*, intended for installation in cases of femurs with osteoarthritis disease [6],
- *Revision*, used for installation as a replacement for primary endoprostheses, [13]
- *Reconstruction*, used for severe bone diseases

Regardless of the categorization, all types of hip endoprostheses have the same geometrical segments, defined by a series of surfaces. These segments have a specific purpose in the endoprosthesis. Fig. 7 shows the structures of the endoprosthesis body with distinctive segments.

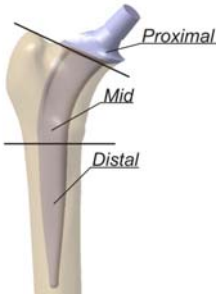


Figure 7
Geometrical segments of the femoral prosthesis body

As it can be seen in Fig. 7, the body of the endoprosthesis has, considering the form and function, three segments:

- *distal*, which allows the positioning of endoprosthesis body in the medullary channel of the femur
- *mid*, segment, which passes through the porous part of the bone and follows the anatomical parameters of the femur,
- *proximal*, which contains an artificial femoral neck and the collar.

Each of these segments is described by one or more surfaces, the shape and dimensions of which depend on several influential factors.

2.3 Geometrical Factors Influencing the Total Hip Endoprosthesis

Reconstruction of the bone geometry, which is the basis for defining the geometrical factors required for the design of hip prostheses, involves the use of several pieces of topological information obtained mostly by using diagnostic devices. In addition to analog X-ray based devices, in medical diagnosis usually digital devices based on tomography methods of recording and computerized tomography (CT) or magnetic resonance imaging (MRI) are used. These digital devices deliver a series of images of diseased limbs cross sections at appropriate planes.

By mathematical processing of digital information represented with the diagnostic image, their conversion is done to the form which allows the formation of CAD models of the femur [18]. These models can be in the form of point clouds or complex volumes obtained by voxelization (replacing pixels as the basic element of the image with voxels) [20].

The basic advantage of reconstruction of the femur on the basis of tomography images is the possibility for precise determination of a large number of characteristic points of the external geometry of the medullary channel of the femur. Geometrical factors influencing the femur, defined on the basis of this information, are used in the design of one or more segments of the endoprosthesis.

The distal segment of the endoprosthesis body is the geometrical form which depends on the shape and dimensions of the medullary channel of the femur. For definition of this segment the following parameters are used:

- The position of the medullary channel isthmus
- Lesser trochanter position
- Layout and dimensions of the medullary channel.

The mid segment of the endoprosthesis body provides its proper positioning, enables the transfer of the load from the pelvic region to the foot and ensures its proper position relative to the rest of the body. The basic geometrical elements that define the shape and dimensions of the middle segment of the endoprosthesis are:

- Position of the lesser trochanter
- Position of the greater trochanter
- The appearance and dimensions of the medullary channel in axial cross-sections 20 mm above and below the lesser trochanter
- The angle of inclination of the body of the femur
- Femoral anteversion angle
- CFI expansion coefficient of the medullary channel

The proximal segment consists of four parts: a body, collar, neck and cone upon which the artificial femoral head is placed. The dimensions and shape of the first two parts are mostly determined for the family of endoprostheses, based on structural and exploitation conditions defined by the manufacturers. The neck and artificial femoral head should ensure proper placement of the body in conjunction with the remaining elements of the endoprosthesis and provide the proper movement of the patient after the hip replacement surgery. The most important geometrical dimensions of the elements that determine these segments are: the position of the femoral head, the distance of the femoral head from the axis, the angle of the femoral neck.

Analysis of hip joint diseases treated by THR method and geometrical parameters necessary to define the existing types of prostheses [7] indicate that the number of influential factors resulting from the femur geometry varies.

2.4 Application of Rational Bezier Curves in Modeling of the Endoprosthesis Body

In computer technology there is a number of mathematical methods that are used to describe complex surfaces. Among them, an important role is played by the rational Bezier curves (Equation 2) belonging to the family of curve surfaces based on the generalized description of the parameter defined polynomial curves.

$$C(t) = \frac{\sum_{i=0}^{n} B_i^n(t) w_i P_i}{\sum_{i=0}^{n} B_i^n(t) w_i} \tag{2}$$

The elements of the equations of these curves are:

- Control points (P_i) which determine the degree of the polynomial and bounds propagation curves
- Elements of the Berstein polynomial $B_i(t)$, described by equation 3 and

$$B_i^n(t) = \binom{n}{i} (1-t)^{n-i} t^i, \; i=0, 1, 2,...n, \; t \in [0,1] \tag{3}$$

- Weighting factors w_i regulating the convergence degree of the curve to control points. Fig. 8 shows the impact of changing weighting factor w2 to the shape of the curve around that point.

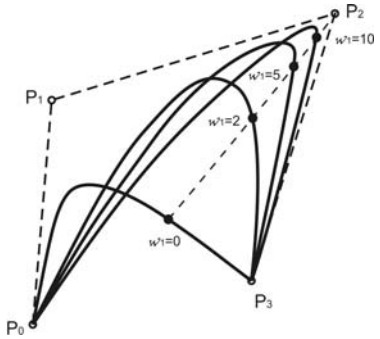


Figure 8
The influence of the weighting factor w_2 to the shape of the rational Bezier curve

Primary advantages of the application of rational Bezier curves in defining the geometry of the computer model are: the ability to correct the curve by using the weighting factor and the possibility of forming an explicit mathematical expression for the curve of a known degree.

By using the same procedure, in addition to spatial curves, we are introducing the concept of rational Bezier surfaces (or the tensor defined surfaces) determined by specific control points (Pi and Pj), by Berstein polynomials Bi (t) and Bj (t), where $t \in [0, 1]$ and the weighting factor wi,j (Equation 4).

$$P(t) = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} B_i^n(t) B_j^m(t) w_{ij} P_i P_j}{\sum_{i=0}^{n} \sum_{j=0}^{m} B_i^n(t) B_j^m(t) w_{ij}} \tag{4}$$

Application of Bezier surfaces allows definition of surface models of objects of arbitrary complexity.

3 General Parametric Model of the Endoprosthesis Body

The general parametric model of the endoprosthesis body is a structure composed of geometrical elements described by functional and discrete parametric models. The main characteristic of the general model is the ability to implement all mentioned influential factors that are necessary for the description of the endoprosthesis body and a possibility to add the new ones. In this specific case, the model is based on the two parts that have a different role in the endoprosthesis. These are:

- the *femoral part* that contains the distal and mid segment of the endoprosthesis is defined by rational Bezier function and

- the *proximal part* that consists of the proximal segment is defined by the three, relatively simple, geometrical shapes.

The femoral part includes segments of the endoprosthesis body (the distal and mid segment) that are located within the femur after the installation. It is specified by a number of parameters, resulting from the geometry of the femur, and a series of exploitation demands arising from the operational method used and the need for increased stiffness of the femur and endoprosthesis body assembly [1, 4] (Fig. 9).

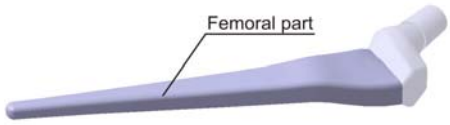


Figure 9 Femoral part of endoprosthesis body

Bezier surface describing the femoral part of endoprosthesis body is formed by combining the two planar Bezier functions in the characteristic planes of the femur (the axial and coronal plane).

In the axial plane, in which the cross-section of the endoprosthesis body is defined, the complexity of the medullary channel and the number of significant factors that are taken into account, are the reason why the curve is determined by Bezier planar function of the tenth degree. Due to better control of the parameters that determine the control points, the curve sections are arranged in the schematic diagram shown in Fig. 10, where they are defined by the distance from the horizontal and vertical axes. This description of the cross-section provides definition of endoprostheses with the symmetric as well as with the asymmetric cross-section.

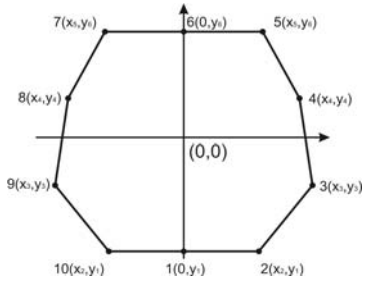


Figure 10
Schematic layout of control points of the endoprosthesis body cross-section

As it is shown on the previous figure, some points are mutually interrelated with the constraints in values of coordinates. This was introduced in order to reduce the degree of uncertainty of the system of equations which determine the parameters of Bezier surfaces, bearing in mind that they are defined on the basis of the influential factors on the body of the endoprosthesis.

The implementation of these points in the curve equation a function of the endoprosthesis cross-section is obtained expressed by a Bezier curve [12].

$$A(t)=\frac{\sum_{i=0}^{10} B_i^{10}(t) w_i P_i}{\sum_{i=0}^{10} B_i^{10}(t) w_i} \tag{5}$$

As noted above, for defining the geometry of the endoprosthesis body in coronal plane along the distal segment, the two characteristic points are necessary (the midpoint of the medullary channel at the narrowest part and 20 mm below the lesser trochanter). Additionally, the mid segment is determined with the three points (the midpoint of medullary channel 20 mm below the lesser trochanter, 20 mm above it and in the plane in which the femoral neck is surgically removed). On the basis of this, it was concluded that the generators of the endoprosthesis can be described by four characteristic points (the plane 20 mm below the lesser trochanter is a place where the distal and middle segments are connecting), so the Bezier curve in the coronal plane can be expressed as follows:

$$B(t)=\frac{\sum_{j=0}^{4} B_j^{4}(t) w_j P_j}{\sum_{j=0}^{4} B_i^{4}(t) w_j} \tag{6}$$

On the basis of the existing partial equations a Bezier surface function can be imported, which includes the aforementioned curves and allows the definition of endoprosthesis segments located within the femur.

$$P(t)=\frac{\sum_{i=0}^{10}\sum_{j=0}^{4} B_{i,p}^{10}(t) B_{j,p}^{4}(t) w_{ij} P_i P_j}{\sum_{i=0}^{10}\sum_{j=0}^{4} B_{i,p}^{10}(t) B_{j,p}^{4}(t) w_{ij}} \tag{7}$$

Fig. 11 shows an example of the femoral endoprosthesis described by the above-mentioned equations.

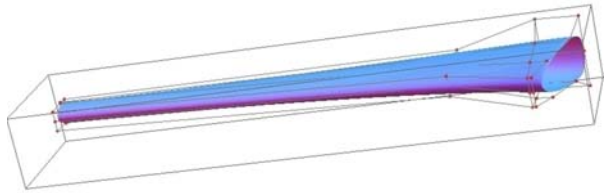


Figure 11 The example of the mathematical description of the femoral part

The proximal segment of the endoprosthesis body is structurally simpler segment of the endoprosthesis body. It contains a series of standardized elements which provide reliance of the endoprosthesis on the femur, and provide the proper formation of the artificial hip joint. These are (Fig. 12):

- the body of proximal segment (T_{ps}) whose geometry depends on the type of endoprosthesis,

- the artificial femoral neck (N_{ps}), which is usually in the form of cone and
- the tapered connection element between the body and endoprosthesis artificial femoral head (usually a cone with slope 1:10 or Morse taper (T_{ps})).

Figure 12
The proximal segment of the endoprosthesis

Because of the simple geometry and a small number of influential factors that determine it, this segment is in the general parametric model implemented as a whole, described by geometrical shapes based on discrete parameters. Fig. 13 shows the final model of the endoprosthesis body obtained by applying a general parametric model.

Figure 13
The hip joint endoprosthesis body

4 Results

In order to verify the general parametric model of the hip endoprosthesis body, a geometrical modeling for different types of endoprostheses that are in clinical use was done.

As the first phase, a verification of the segment of the general parametric model, used to define the geometry of the intersection of the endoprosthesis was carried out. This was realized by defining several characteristic cross-sections that are used for the body of the endoprosthesis. Fig. 14 shows the cross-sections of the most commonly used types of hip joint endoprosthesis body [15].

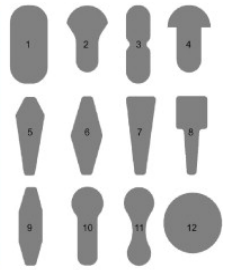


Figure 14
Cross-sections of different endoprosthesis body types [15]

The research [15] has indicated that, from those shown, the most common cross-sections used in practice include: circular (especially in tumors and total prostheses), trapezoidal and combined shape.

Based on the analysis of the frequency of their application, the four endoprosthesis profiles were chosen: round, rectangular, trapezoidal, and combined ("mushroom like"), and curves were defined for them based on real dimensions of the cross-section of the femur. Fig. 15 a, b, c and d shows the cross-sections of the endoprosthesis body described by Bezier functions.

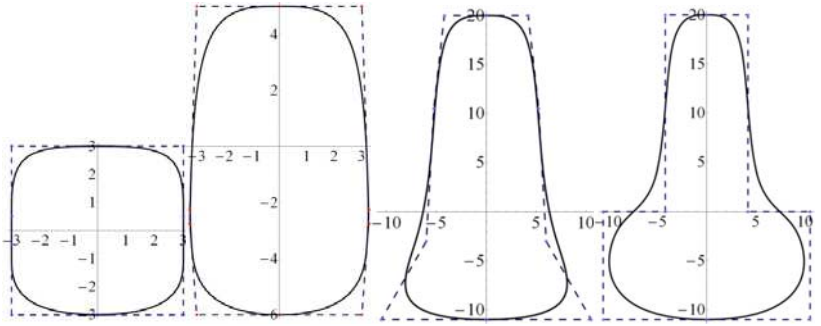


Figure 15
Mathematical interpretation of the endoprosthesis body cross-section

Table 1 shows the list of matrices containing the coordinates of control points for real dimensions of the analyzed sections.

Table 1
Coordinates of control points

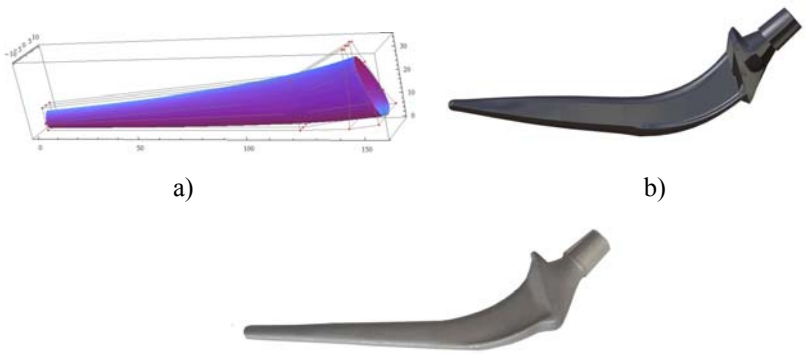
Type of curve	Matrix of coordinates of points
Cross-section a	$\begin{bmatrix} 0 & 3 & 3 & 3 & 3 & 0 & -3 & -3 \\ 3 & -3 & -.5 & .5 & 3 & 3 & 3 & .5 \end{bmatrix}$
Cross-section b	$\begin{bmatrix} 0 & 3 & 3.25 & 3.25 & 3 & 0 & -3 & -3.25 & -3.25 & -3 \\ -6 & -6 & -2.75 & 2.25 & 5 & 5 & 5 & -2.25 & -2.75 & -6 \end{bmatrix}$
Cross-section c	$\begin{bmatrix} 0 & 10.5 & 6 & 5.25 & 4.2 & 0 & -4.2 & -5.25 & -6 & -10.5 \\ -11 & -11 & -3 & 10.5 & 20 & 20 & 20 & -10.5 & -3 & -11 \end{bmatrix}$
Cross-section d	$\begin{bmatrix} 0 & 10.5 & 10.5 & 4.2 & 4.2 & 0 & -4.2 & -4.2 & -10.5 & -10.5 \\ -11 & -11 & 0 & 0 & 20 & 20 & 20 & 0 & 0 & -11 \end{bmatrix}$

To reproduce these curves it is important to note that here the weighting factor 1 (one) was used for each characteristic point. By increasing the weighting factors the profile can be adjusted to the control points.

In the second stage of verification, the endoprosthesis body modeling for the three types of clinically tested, and for several years exploited, endoprostheses was done. These are:

- The body of the primary hip endoprosthesis BB2 (manufactured by DES Novi Sad – Serbia);
- The body of anatomical endoprosthesis BB3 (developed by the Faculty of Technical Sciences in Novi Sad in cooperation with the Institute of Orthopedic Surgery Banjica, Serbia) with the anteversion in the middle part;
- Body of the tumor endoprosthesis (manufacturer Grujic & Grujic, Serbia).

For verification purposes, and the further studies, the general parametric model is implemented in a software system CATIA. Modeling was realized in two phases where the spatial surfaces of the femoral part were formed, which, after conversion into a volume, were integrated with the proximal segment of the endoprosthesis. Fig. 16 shows a model of the femoral endoprosthesis BB2 defined by Bezier curve surfaces in Wolfram Mathematica software system (Fig. 16a), the same model defined in CATIA software system with the additional ribs (Fig. 16b), as well as photo of the body of endoprosthesis to which an element in the form of ribs was added to increase rigidity (Fig. 16c).

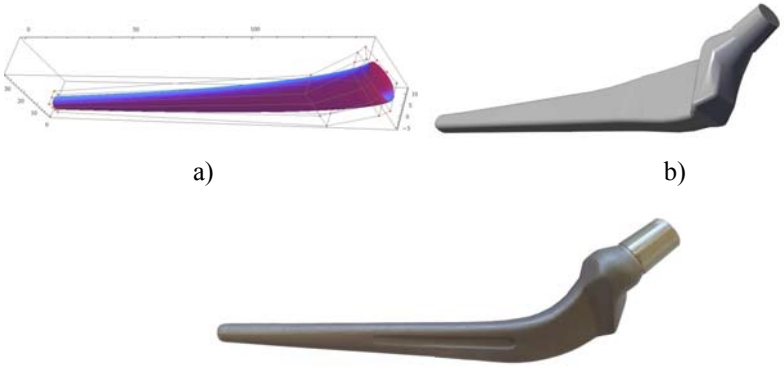


a) b)

c)

Figure 16
Mathematical and CAD model and a photograph of the femoral endoprosthesis BB2

The second type of endoprosthesis which was modeled is geometrically more complex and belongs to the group of so-called anatomical endoprostheses containing anteversion. From this point of view this type of endoprosthesis is more complex for modeling. Fig. 17 presents a model of the femoral endoprosthesis BB3 defined by Bezier curve surfaces in Wolfram Mathematica software system (Fig. 17a), its interpretation in the CATIA software system (17b) and a photo of the endoprosthesis (17c).



a) b)

c)

Figure 17
Mathematical and CAD model and a photograph of the femoral endoprosthesis BB3

Finally, the last type of endoprosthesis which was modeled was a total endoprosthesis of the hip joint, which is used to treat tumors (sarcomas)[19]. This type of endoprosthesis is geometrically the simplest because it consists of mostly

conical elements. This means that the femoral part is defined as a cone that is smaller in length than other types of prostheses, since the surgery removes most of the femur. Fig. 18 presents the femoral part, as described in the Mathematica software system (18a) and CATIA (18b) and a corresponding photo (18c).

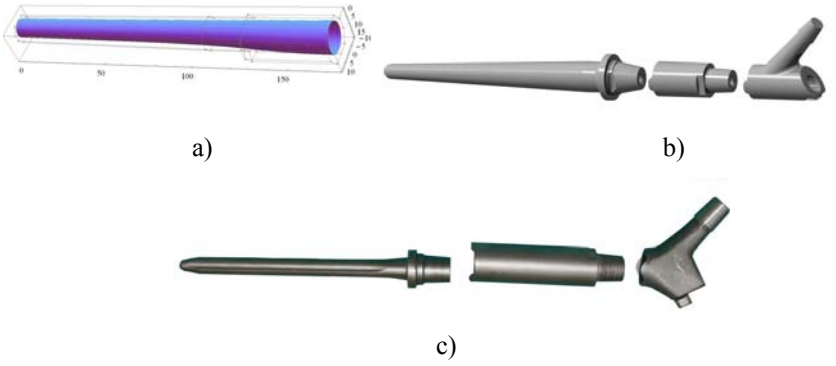


Figure 18
Mathematical and CAD model and a photograph of the femoral part of tumor endoprosthesis BB3

5 Discussion

The research results presented in this paper can be observed from the two perspectives: the analysis of the applicability of the general parametric model of the endoprosthesis body, and through the results of verification.

Based on the description and interpretation of the structure of the general parametric model (especially based on the Fig. 15), it can be concluded that the presented model can be applied only in the design of prostheses which is symmetric with one of the axes in the cross section plane. This is a consequence of the aim to minimize the number of variables that occur in the design of endoprostheses. On the other hand, the analysis of current endoprostheses solutions suggests that in clinical practice endoprostheses with the asymmetrical cross section are not used. In this way, a mathematical interpretation of the endoprosthesis body model is obtained so that approximately the same number of geometrical parameters (point coordinates) is used as in the case of models with discretely defined parameters. The basic advantage of this model is significantly increased flexibility, which is reflected in the possibility to adapt the shape of the endoprosthesis to the medullary channel of the femur, as well as introduction of one generic model for all standardized forms of endoprosthesis body.

Verification presented in this paper was carried out to confirm the ability of the general geometrical model to conform to the shape and profile of the existing forms of enodprostheses. Since the aim of introducing a general model of the design of this product is to promote the development of custom-made prostheses that can be adjusted to the specific femur, a more detailed verification with approximation of errors of described shapes is not applicable here. Further improvement of the general model should include an analysis of the possibilities to approximate forms of medullary channel while taking into consideration other influential factors.

Conclusions

Designing the hip joint endoprostheses has, for many years, been one of the interesting areas of scientific research activities, due to the large number of influencing factors and complex geometry.

Available results indicate that there are two directions of endoprosthesis design improvement. The first direction is related to a research aimed at improving the characteristics of the endoprosthesis by introducing new influential factors obtained by simulation of the behavior of the endoprosthesis body as a result of FEA analysis, experiments and clinical monitoring of recovery process in patients. The second direction is related to the research targeted at improving the design process to allow creation of the endoprosthesis with measures adjusted to a specific patient, over a short period of time and at a reasonable price. From the engineering point of view these two contradictory demands can be met by increasing the flexibility of computer models and with their structural simplification, based on which the development of CAD systems with the partial automation of the design process becomes possible.

This paper describes the structure of an original general parametric model of the hip endoprosthesis body, adapted to the observed endoprosthesis improvement directions. The main advantages of general parametric model usage are the increased flexibility of the model as well as the model structure that allows calculation of the geometrical parameters independently of the CAD program used to realize the modeling. In this way, a basis for a fully efficient practical automation of some phases of the endoprosthesis design process is created.

References:

[1] P. Benum, A. Aamodt, “Uncemented Custom Femoral Components in Hip Arthroplasty A Prospective Clinical Study of 191 Hips Followed for at Least 7 Years”, *Acta Orthopaedica*, Vol. 81, pp. 427-435, 2010

[2] J. J. Callaghan, A. G. Rosenberg, H. E. Rubash, “The Adult Hip”, Lippincott Williams & Wilkins, 1997

[3] P. Cerveri, M. Marchente, W. Bartels, K. Corten, J. P. Simon, A. Manzotti, “Automated Method for Computing the Morphological and Clinical Parameters of the Proximal Femur Using Heuristic Modeling Techniques”,

Annals of Biomedical Engineering, Vol. 38, pp. 1752-1766, 2010

[4] R. D. Crowninshield, R. A. Brand, R. C. Johnston, J. C. Milroy, "An Analysis of Femoral Component Stem Design in Total Hip Arthroplasty", *The Journal of bone and joint surgery. American volume*, Vol. 62, pp. 68-78, 1980

[5] X. Dong, G. Zheng, "Fully Automatic Determination of Morphological Parameters of Proximal Femur from calibrated Fluoroscopic Images through Particle Filtering", Proceedings of the Third international conference on Image Analysis and Recognition, Vol. II, Springer-Verlag, Portugal, pp. 535-546, 2006

[6] N. P. Hailer, G. Garellick, J. Karrholm, "Uncemented and Cemented Primary Total Hip Arthroplasty in the Swedish Hip Arthroplasty Register Evaluation of 170,413 operations", *Acta Orthopaedica*, Vol. 81, pp. 34-41, 2010

[7] ISO 7206-1-2008, "Implants for Surgery — Partial and Total Hip Joint Prostheses — Part 1: Classification and Designation of Dimensions", *International Organization for Standardization*, Geneva, Switzerland, 2008

[8] Y. Jun, K. Choi, "Design of Patient-Specific Hip Implants Based on the 3D Geometry of the Human Femur", *Advances in Engineering Software*, Vol. 41, pp. 537-547, 2010

[9] Y. Jun, "Morphological Analysis of the Human Knee Joint for Creating Custom-Made Implant Models", *International Journal of Advanced Manufacturing Technology*, Vol. 52, pp. 841-853, 2011

[10] Y. Kalairajah, S. Molloy, M. Patterson, "The Effect of Femoral Stem Size on Failure Rates in Total Hip Replacement Cemented with Boneloc", *Acta orthopaedica Belgica*, Vol. 68, pp. 33-36, 2002

[11] K. Kawate, Y. Ohneda, T. Ohmura, H. Yajima, K. Sugimoto, Y. Takakura, "Computed Tomography-based Custom-Made Stem for Dysplastic Hips in Japanese Patients", *Journal of Arthroplasty*, Vol. 24, pp. 65-70, 2009

[12] H. J. Laine, M. Lehto, T. Moilanen, "Diversity of Proximal Femoral Medullary Canal", *Journal of Arthroplasty*, Vol. 15, pp. 86-92, 2000

[13] H. Malchau, L. Ahnfelt, "Prognosis of Total Hip Replacement in Sweden: Follow-Up of 92,675 Operations Performed 1978-1990", Acta Orthopaedica Scandinavica, Vol. 64, pp. 497-506, 1993

[14] D. Patel, T. Goswami, "Influence of Design Parameters on Cup-Stem Orientations for Impingement Free RoM in Hip Implants", *Medical Engineering & Physics*, Vol. 34, pp. 573-578, 2012

[15] A. Ramos, A. Completo, C. Relvas, J. A. Simoes, "Design Process of a Novel Cemented Hip Femoral Stem Concept", *Materials & Design*, Vol. 33, pp. 313-321, 2012

[16] B. R. Rawal, R. Ribeiro, R. Malhotra, N. Bhatnagar, "Design and Manufacturing of Femoral Stems for the Indian Population", *Journal of Manufacturing Processes*, Vol. 14, pp. 216-223, 2012

[17] B. B. Seedhom, N. C. Wallbridge, "Walking Activities and Wear of Prostheses", *Annals of the Rheumatic Diseases*, Vol. 44, pp. 838-843, 1985

[18] W. Sun, B. Starly, J. Nam, A. Darling, "Bio-CAD Modeling and its Applications in Computer-aided Tissue Engineering", *Computer-Aided Design*, Vol. 37, pp. 1097-1114, 2005

[19] S. Tabakovic, A. Zivkovic, J. Grujic, M. Zeljkovic, "Using CAD/CAE Software Systems in the Design Process of Modular, Revision Total Hip Endoprosthesis", *Academic Journal of Manufacturing Engineering – AJME*, Vol. 9, pp. 97-102, 2011

[20] D. J. Yoo, "Three-Dimensional Surface Reconstruction of Human Bone Using a B-Spline-based Interpolation Approach", *Computer-Aided Design*, Vol. 43 pp. 934-947, 2011