

Grid-based Support for Different Text Mining Tasks

Martin Sarnovský, Peter Butka, Ján Paralič

Centre for Information Technologies
Department of Cybernetics and Artificial Intelligence
Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9, 04200 Košice, Slovakia
E-mail: martin.sarnovsky@tuke.sk, peter.butka@tuke.sk, jan.paralic@tuke.sk

Abstract: This paper provides an overview of our research activities aimed at efficient use of Grid infrastructure to solve various text mining tasks. Grid-enabling of various text mining tasks was mainly driven by increasing volume of processed data. Utilizing the Grid services approach therefore enables to perform various text mining scenarios and also open ways to design distributed modifications of existing methods. Especially, some parts of mining process can significantly benefit from decomposition paradigm, in particular in this study we present our approach to data-driven decomposition of decision tree building algorithm, clustering algorithm based on self-organizing maps and its application in conceptual model building task using the FCA-based algorithm. Work presented in this paper is rather to be considered as a 'proof of concept' for design and implementation of decomposition methods as we performed the experiments mostly on standard textual databases.

Keywords: Grid services, text mining, clustering, classification, formal concept analysis

1 Introduction

The process of knowledge discovery is one of the most important topics in scientific and business problems. Nowadays, when the information overload means a big problem, knowledge discovery algorithms applied on very large text document collections can help to solve numerous problems and as text is still premier source of information on the web, the role of text mining is increasing. However, data are often geographically distributed in various locations. One approach to face this problem is distributed computing - distributed text mining algorithms can offer an effective way to mine extremely large document collections.

Motivation of this work is to use the Grid computational capabilities to solve text mining tasks. Grid is a technology, that allows from geographically distributed computational and memory resources create a universal computing system with extreme performance and capacity [1]. Nowadays the Grid projects are built on protocols and services that enable applications to handle distributed computing resources as a single virtual machine.

Some of the methods are time-consuming and use of the Grid infrastructure can bring significant benefits. Implementation of text mining techniques in distributed environment allows us to perform text mining tasks, such as text classification, in parallel/distributed fashion.

Knowledge discovery in texts is a variation of a field called knowledge discovery in databases, that tries to find interesting patterns in data. It is a process of semiautomatic non-trivial extraction of previously unknown, potentially useful and non-explicit information from large textual document collection. A key element of text mining is to link extracted information together to form new facts or new hypotheses to be explored further by more conventional means of experimentation. While regular data mining extracts the patterns from structured databases, text mining deals with problem of natural language processing. The biggest difference between data mining and text mining is in the preprocessing phase. Preprocessing of text documents is completely different than in the case of databases; in general, it is necessary to find a suitable way to transform the text into an appropriate internal representation, which the mining algorithms can work with. One of the most common internal representations of document collections is Vector Space Model [2]. Text mining phase is the core process of knowledge discovery in text documents. There are several types of text mining tasks as follows:

- Text categorization: assigning the documents with pre-defined categories (e.g. decision trees induction).
- Text clustering: descriptive activity, which groups similar documents together (e.g. self-organizing maps).
- Concept mining: modelling and discovering of concepts, sometimes combines categorization and clustering approaches with concept/logic-based ideas in order to find concepts and their relations from text collections (e.g. formal concept analysis approach for building of concept hierarchy).
- Information retrieval: retrieving the documents relevant to the user's query.
- Information extraction: question answering.

It is very usual that in any text mining process first three types are basic elements in order to support also information retrieval/extraction. Main goal of our work is

to show how well-known methods for text categorization, text clustering and concept mining could be adopted in Grid (distributed) environment in order to achieve more robust and faster application of text mining tasks. In our case we have implemented and tested three candidates, one per each type. After briefly presenting related work in Section 2, we describe every method with all modifications we have used in Section 3. In Section 4 we provide proposal and implementation details of Grid-based support in every case, and describe our experiments, results achieved and their evaluation (with emphasis on distribution aspects) in Section 5. Finally, we sum up the main results in conclusions section.

2 Related Work

In this section, we briefly describe related projects that utilize the Grid to perform advanced knowledge discovery in textual documents. DiscoveryNet¹ provides a service-oriented computing model for knowledge discovery, allowing the user to connect to and use data analysis software as well as document collection that are made available online by third parties. The aim of this project is to develop a unified real-time e-Science text mining infrastructure that leverages the technologies and methods developed by the DiscoveryNet and myGrid² projects. Both projects have already developed complementary methods that enable the analysis and mining of information extracted from biomedical text data sources using Grid infrastructures, with myGrid developing methods based on linguistic analysis and DiscoveryNet developing methods based on data mining and statistical analysis. National Centre for Text Mining³ is also involved in research activities covering the Grid based text mining. Primary goal of this project is also focused to develop an infrastructure for text mining, a framework comprised of high-performance database systems, text and data mining tools, and parallel computing. Our work, presented in this article is complementary to the previous projects. Some of our algorithms (classification and clustering tasks) have been used within the GridMiner project⁴. Moreover, the FCA approach as far as we know has not been approached in any of the projects listed above.

¹ www.discovery-on-the.net

² www.myGrid.org.uk

³ www.cse.salford.ac.uk/nactem

⁴ www.gridminer.org

3 Text Mining Algorithms

3.1 Classification Using Decision Trees

Text Classification is the problem of assigning a text document into one or more topic categories or classes based on document's content. Traditional approaches to classification problems usually consider only the uni-label classification problem. It means that each document in collection has associated one unique class label. This approach is typical for data mining classification tasks, but in a number of real-world text mining applications, we face the problem of assigning the document into more than one single category. One sample can be labelled with a set of classes, so techniques for the multi-label classification problem have to be explored. Especially in text mining tasks, it is likely that data belongs to multiple classes, for example in context of medical diagnosis, a disease may belong to multiple categories, genes may have multiple functions, etc. In general there are many ways to solve this problem. One approach is to use a multinomial classifier such as the Naive Bayes probabilistic classifier that is able to handle multi-class data. But most of commonly used classifiers (including decision trees) cannot handle multi-class data, so some modifications are needed. Most frequently used approach to deal with multi-label classification problem is to treat each category as a separate binary classification problem, which involves learning a number of different binary classifiers and use an output of these binary classifiers to determine the labels of a new example. In other words, each such problem answers the question, whether a document should be assigned to a particular class or not. In the work reported in this paper, we used the decision trees algorithm based on the Quinlan's C4.5 [3]. A decision tree classifier is a tree with internal nodes labelled by attributes (words), branches departing from them are labelled by tests on the weight that attribute has in the document, and leafs represent the categories [4]. Decision tree classifies the unknown example by recursively testing of weights in the internal nodes, until a leaf is reached. While this algorithm is not suitable to perform multi-label classification itself, we use the approach of constructing different binary tree for each category. The process of building many binary trees can be very time consuming when running sequentially, especially on huge document collections. Due to the fact that these binary classifiers are independent on each other, it is natural to find a suitable way how to parallelize the whole process. Growing of these binary trees is ideal for parallel execution on a set of distributed computing devices. Such a distribution might be desirable for extremely large textual document collections or large number of categories, which e.g. can be associated with a large number of binary classifiers.

3.2 Clustering Using Growing Hierarchical Self-Organizing Maps

Self-organizing maps (SOM) [5] algorithm is one of the methods for non-hierarchical clustering of objects based on the principles of unsupervised competitive learning paradigm. This model provides mapping from high-dimensional feature space into (usually) two-dimensional output space called map, which consists of neurons characterized by n -dimensional weight vector (same dimension as input vectors of the objects). Specific feature of SOM-based algorithms is realization of topology preserving mapping. Neurons are ordered in some regular structure (e.g. usually it is simple two-dimensional Grid) representing output space. A distance measure used in this space could be e.g. Euclidean distance based on the coordinates of weight vectors of neurons in the output space. Mapping created by learning of SOM then has a feature that two vectors which are closed each other in the input space are also mapped onto closely located neurons in the output space. Training consists of two steps: presentation of input document at the network input and adaptation of weight vectors. Based on the activation of the network best candidate from the neurons on the map is used as winner (e.g. Euclidean distance is used to find lowest distance to input vector). Next, weight vector of the winner and its neighbourhood neurons (with descending influence) are adapted in order to decrease its distance to the input vector.

One of the disadvantages of SOM algorithm is that structure of the whole output space is defined apriori. It is possible to avoid these using modifications, which dynamically expand map according to needs of the input feature space. The problem of adapting map is that it could expand to really large Grid (so we get same result like in case of SOM structure with predefined larger size) and in some applications it is hard to interpret results usefully. This leads us to go for another modification in “hierarchical” dimension – algorithm called GHSOM (Growing Hierarchical Self Organizing Map) [6], where map expands in two different ways:

- Hierarchically – according to the data distribution of input vectors some neurons on a map with large number of input documents assigned to them should be independently clustered in separate maps (each of these neurons expands into a submap on lower level), this provides hierarchical decomposition and navigation in submaps (Hierarchical SOM part).
- Horizontally – change of size of particular (sub)maps according to requirements of the input space (as it is done by Growing SOM).

Algorithm GHSOM consists of these steps:

- 1 First, *mean quantization error* (deviation of all input vectors) is computed on at layer 0 (it could be seen as mean deviation of all input vectors with respect to a map with just one neuron - cluster). Then weight

vector $m_0 = [\eta_{01}, \eta_{02}, \dots, \eta_{0n}]^T$ contains average values for every attribute from the whole collection of input vectors. Mean quantization error of layer 0 is simply:

$$mqe_0 = \frac{1}{d} \cdot \|m_0 - x\|,$$

where d is the number of input vectors x .

- 2 Learning of GHSOM starts with the layer on level 1. This map is usually small (e.g. 2x2). For every neuron i we need n -dimensional weight vector

$$m_i = [\eta_{i1}, \eta_{i2}, \dots, \eta_{in}]^T, m_i \in \mathfrak{R}^n,$$

which is initialized randomly. Dimension n of these vectors has to be the same as dimension of input vectors.

- 3 Learning of SOM is competitive process between neurons for better approximation of input vectors. Neuron with weight vector nearest to the input vector is the winner. Its weight vector as well as weight vectors of neurons in its neighborhood is adapted in order to decrease their difference to input vector. The grade of adaptation is controlled by learning parameter $\alpha(t)$, which is decreasing during time of learning. Number of neighboring neurons, which are also adapted, is also decreasing with time. At the beginning of the learning process many of winner's neighbors are adapting, but near the end of learning only the winner is adapted. Which neurons and how much are adapted is defined by the neighborhood function $h_{ci}(t)$, which is based on distance between winner c and current neuron i (in output space). As a combination of these principles we have the following learning rule for computing of weight vector m_i :

$$m_i(t+1) = m_i(t) \cdot \alpha(t) \cdot h_{ci}(t) \cdot [x(t) - m_i(t)],$$

where x is actual input vector, i is current neuron and c is winner in iteration t .

- 4 After some number of iterations (parameter λ) mean quantization error of map is computed using:

$$MQE_m = \frac{1}{u} \cdot \sum_i mqe_i,$$

where u is number of neurons i at map m , mqe_i is mean quantization error of neuron i at the map m . Every layer of the GHSOM is responsible for explaining some portion of the deviation of the input data as present in its preceding layer. This could be achieved by adding of new neurons into map on every layer in order to have suitable size. Maps on every level grow until the deviation

present in the unit of its preceding layer is reduced to at least a fixed percentage τ_m . The smaller the parameter τ_m is chosen, the larger will be the size of SOM. If for current map condition

$$MQE_m \geq \tau_m \cdot mqe_0$$

is fulfilled, new row or column of neurons is added into map. It is added near the error neuron (neuron with largest error). Addition of row or column depends on position of most distant neighbor neuron to error neuron (new row or column is inserted between them; distance is computed in input space – weight vectors of neurons). Weight vectors of new neurons are usually initialized as average values of neighboring neurons. After such a neuron addition learning parameters are setup to starting values and map is re-learned.

- 5 When the learning of map on level 1 (or any other level) is finished, it means that

$$MQE_m < \tau_m \cdot mqe_0,$$

it is a time to expand neurons of the map to another level (if needed). Neurons, which have still high mean quantization error (comparing with mqe_0), should be expanded and new map in next hierarchical level is created. Every neuron i , which fulfils next condition, have to be expanded:

$$mqe_i > \tau_u \cdot mqe_0,$$

where τ_u is parameter for controlling of hierarchical expansion.

- 6 Learning process follows for every new map identically with steps 2 to 5. Only difference is that in every new submap only inputs from one expanded neuron of parent map are used for learning of its submap, and only fraction of quantization error of the parent map is going to be analyzed (concretely error of expanded neuron).
- 7 GHSOM algorithm is finished when there is no neuron for expansion, or some predefined maximal depth of hierarchy is reached.

To summarize, the growth process of the GHSOM is guided by just two parameters. The parameter τ_u specifies the desired quality of input data representation at the end of the training process in order to explain the input data in more detail (if needed). Contrary to that, the parameter τ_m specifies the desired level of detail that is to be shown in one SOM. Hence, the smaller τ_m the larger will be the emerging maps. Conversely, the larger τ_m the deeper will be the hierarchy.

3.3 Use of Formal Concept Analysis in Text Analysis

Formal Concept Analysis (FCA, [7]) is a theory of data analysis that identifies conceptual structures among data sets. FCA is able to identify and describe all concepts (extensionally) and discover their structure in form of conceptual lattice that can be e.g. graphically visualized. These formal structures present inherent structures among data that (in our case) can be understood as knowledge model – e.g. ontology. It is an explorative method for data analysis and provides nontrivial information about input data of two basic types – concept lattice and attribute implications. Concept is cluster of “similar” objects (similarity is based on presence of the same attributes); concepts are hierarchically organized (specific vs. general). Standard usage of FCA is based on binary data tables (object has/has not attribute) – crisp case.

Problem is that classic data table from textual documents contains real-valued attributes. Then we need some fuzzification of classic crisp method. One approach to one-sided fuzzification was presented in [8]. Concept lattice created from real-valued (fuzzy) attributes is called *one-sided fuzzy concept lattice*. The proposed algorithm for FCA discovery is computationally very expensive and provides a huge amount of concepts (if we use definition). One approach to solve the issue is based on the problem decomposition method (as was described in [9]). This paper describes one simple approach to creation of simple hierarchy of concept lattices. Starting set of documents is decomposed to smaller sets of similar documents with the use of clustering algorithm. Then particular concept lattices are built upon every cluster using FCA method and these FCA-based models are combined to simple hierarchy of concept lattices using agglomerative clustering algorithm. For our experiments we used GHSOM algorithm for finding of appropriate clusters, then ‘Upper Neighbors’ FCA algorithm (as defined in [10]) was used for building of particular concept lattices. Finally, particular FCA models were labelled by some characteristic terms and simple agglomerative algorithm was used for clustering of local models, with the metric based on these characteristic lattices terms. This approach is easy to implement in distributed manner, where computing of local models can be distributed between nodes and then combined together. This will lead to reduction of time needed for building the concept model on one computer (sequential run). Next, we will shortly describe the idea of one-sided fuzzy concept lattice, process of data pre-clustering, concept lattices creation and algorithm for combination of concept lattices (introduced in [18]).

3.3.1 One-sided Fuzzy Concept Lattice

Let A (attributes) and B (objects) are non-empty sets and R is fuzzy relation on their Cartesian product, $R: A \times B \rightarrow [0,1]$. This relation represents real-valued table data with rows and columns as objects and attributes, respectively. In case of texts, object is document and attribute is term (word). Then $R(b,a)$ express a grade

in which the document b contains term a (or in text mining terminology – weight of term a in vector representation of document b).

Now we can define mapping $\tau: P(B) \rightarrow \hat{[0,1]}$ which assigns to every set X of elements of B function $\tau(X)$ with value in point $a \in A$ (P – power set):

$$\tau(X)(a) = \min\{R(a,b) : b \in X\},$$

i.e. this function assigns to every attribute the least of such values. This means that objects from X have this attribute at least in such grade.

Another (backward) mapping $\sigma: \hat{[0,1]} \rightarrow P(B)$ then simply assigns to every function $f: A \rightarrow [0,1]$ a set:

$$\sigma(f) = \{b \in B : (\forall a \in A) R(a,b) \geq f(a)\},$$

i.e. those attributes, which have all values at least in grade set by the function f (these attributes the function of their fuzzy-membership to objects dominates over f). From properties of mappings we can see that the pair $\langle \tau, \sigma \rangle$ is Galois connection, i.e. $\forall X \subseteq B$ and $f \in \hat{[0,1]}$ holds $f \leq \tau(X)$ iff $X \subseteq \sigma(f)$.

Now we can define mapping $cl: P(B) \rightarrow P(B)$ as the composition of the mappings τ and σ , i.e. $\forall X \subseteq B : cl(X) = \sigma(\tau(X))$. Because conditions $X \subseteq cl(X)$, $X_1 \subseteq X_2 \rightarrow cl(X_1) \subseteq cl(X_2)$ and $cl(X) \subseteq cl(cl(X))$ are fulfilled, cl is a closure operator.

As in crisp case, concepts are subsets $X \subseteq B$ for which $X = cl(X)$. Such pair $\langle X, \tau(X) \rangle$ is called *(one-sided) fuzzy concept* (X – extent of this concept, $\tau(X)$ – intent of this concept). Then the set of all concepts, i.e. $L = \{X \in P(B) : X = cl(X)\}$, ordered by inclusion is a lattice called *one-sided fuzzy concept lattice*, operation of which are defined as following: $X_1 \wedge X_2 = X_1 \cap X_2$ and $X_1 \vee X_2 = cl(X_1 \cup X_2)$. In next parts we will use only term concept lattice, but we mean always one-sided fuzzy concept lattice presented above.

3.3.2 Proposed Approach for Using of Problem Decomposition Method

In our case, FCA can be used to create hierarchy of concepts and relations between these concepts. Problems with use of this method in textual documents domain is time-consuming computation of concepts and hard interpretability of huge amounts of concepts. Solution can be combination with other algorithms like clustering algorithms. Problem decomposition approach can be seen on Fig. 1 [9].

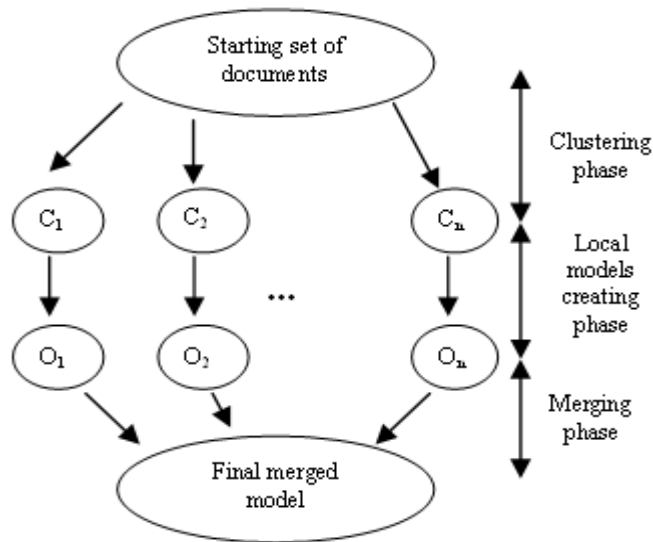


Figure 1

This diagram presents general scheme of the reduction-based conceptual model creation step. In clustering phase input dataset is divided in many smaller training sets. Then particular model O_i using FCA is created from every cluster C_i . Finally, all local models are merged together in the last phase of model generation step.

“Pre-clustering” of input set of documents can be viewed as reduction step where interesting groups of similar documents are found. The reduction step is based on filtering of terms that these objects (inside cluster) do not contain cooperatively. This step is top-down reduction problem (divide-and-conquer) approach to conceptual model extraction phase. Every cluster has independent training set (with reduced cardinality of weight vector), for each one a small concept lattice is built with a help of fuzzy FCA approach. Small models are merged then together and whole conceptual model from tested collection is finally created.

Important steps of our implementation are (more detailed description of every step is provided in [18]):

- 1 We use GHSOM clustering method for dividing the initial large set of documents into a hierarchy of clusters.
- 2 Find local concept lattices for every cluster of similar documents in the resulted GHSOM ‘leaves’ (neuron without sub-map, in the end of expansion), i.e. in created hierarchy of maps particular one-sided concept lattices are built upon documents using ‘Upper Neighbors’ algorithm (as presented in [10], updated for real-valued attributes). Before creation of the whole concept lattice documents are tested through attributes, if value of some attribute is lower than some threshold, value of attribute is set to zero.

This is inspired by work presented in [11] and is very useful for reduction of number of terms in concept lattice description. If we have higher concept in hierarchy of lattices, the number of concept's terms and weights is smaller. Terms with non-zero weights can be used as characteristic terms of actual concept (set of documents in concept).

- 3 Every concept lattice then can be presented as hierarchy of concepts characterized by some terms. Because we needed some description of lattice for merging of lattice to one model, we extracted terms from particular lattices and created their representation based on these terms. A weight of descriptive terms was based on level of terms in hierarchy (of course, important was highest occurrence of term). Then terms can be used for characterization of particular lattices and for clustering based on some metric.
- 4 Merging phase is based on clustering of lattices. First, we created one node for every local hierarchy (for every concept lattice), which contains list of documents, list of characteristic terms (sorted by value of weights), vector of terms weight's values (also in normalized type). Particular nodes are then compared using vectors of terms' weights, so vectors are normalized into interval $<0,1>$. After this step differences between numbers of documents in particular nodes are respected. Comparison of lattices is used in process of agglomerative clustering of these nodes (for detailed description of algorithm see [18]).

Final hierarchy contains nodes with list of documents in it and the sorted list of characteristic terms of nodes. Every node has link to upper node and list of lower nodes. 'Leaf' nodes of hierarchy contain link on the particular local concept lattices.

4 Distributed Support for Text Mining Algorithms

4.1 Tools and Technologies for Grid-based Support

JBOWL - (Java Bag-of-Words Library) [12] is an original software system developed in Java to support information retrieval and text mining. The system is being developed as open source with the intention to provide an easy extensible, modular framework for pre-processing, indexing and further exploration of large text collections, as well as for creation and evaluation of supervised and unsupervised text mining models. JBOWL supports the document preprocessing, building the text mining model and evaluation of the model. It provides a set of classes and interfaces that enable integration of various classifiers. JBOWL

distinguishes between text mining algorithms (SVM, SOM, linear perceptron) and text mining models (rule based classifiers, classification trees, maps, etc.).

GridMiner [13] is a framework for implementing data mining services in the Grid environment. It provides three layered architecture utilizing a set of services and web applications to support all phases of data mining process. The system provides a graphical user interface that hides the complexity of the Grid, but still offers the possibility to interfere with the data mining process, control the tasks and visualize the results. GridMiner is being developed on top of the Globus Toolkit.

4.2 Distributed Trees Induction

The interface of the sequential and distributed versions of the service defines two main methods needed to build final model: *BuildTextModel* and *BuildClassificationModel*. While the first one is implemented as a pure sequential method, the second one can build the final model distributing the partial binary classifiers [14, 15]. This behaviour of the service depends on its configuration. Moreover, other methods were implemented to provide term reduction and model evaluation, but these methods were not used during the performance evaluation experiments discussed in the next section.

1 *BuildTextModel* - This method creates the Text Model from the documents in the collection. The model contains a document-term matrix created using TF-IDF weighting, which interprets local and global aspects of the terms in collection. The input of the method is a parameter specifying the text model properties and the location of the input collection.

2 *BuildClassificationModel* - The Classification Model, as the result of the decision tree classifier, is a set of decision trees or decision rules for each category. This service method creates such a model from the document-term matrix created in the previous method. The sequential version builds the model for all categories and stores it in one file. The process of building the model iterates over a list of categories and for each of them creates a binary decision tree (based on tree algorithm described in Chapter 3). The distributed version performs the same, but it distributes the work of building individual trees onto other services, so called workers, where partial models containing only trees of dedicated categories are created. These partial models are collected and merged into the final classification model by the master node and stored in the binary file, which can be passed to a visualization service.

4.3 Distributive Approach in Learning of GHSOM Model

Distributed algorithm GHSOM is implemented as service in GridMiner system Grid layer using Jbowl (Java 1.5). Implementation contains distributed algorithm together with preprocessing of text documents and visualization of output model. After creation of first map several new clustering processes are started – building of hierarchical sub-GHSOMs, which consist of hierarchically ordered maps of Growing SOM. Main idea is parallel execution of these clustering processes on working nodes of Grid. Approach can be described easily (scheme of distribution is shown on Fig. 2) [16]:

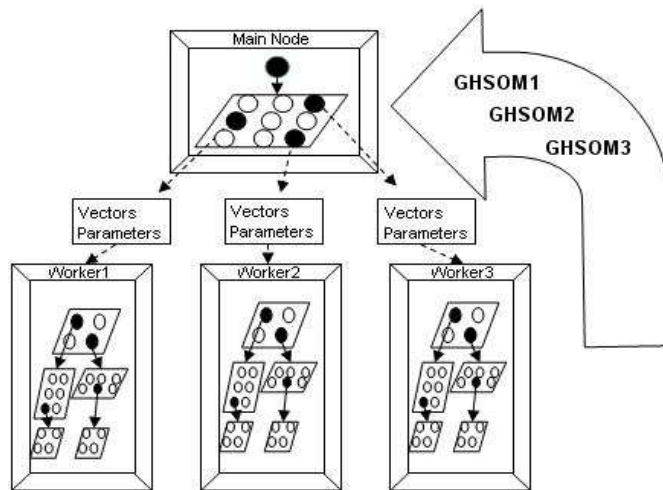


Figure 2

Distribution scheme of GHSOM algorithm in Grid environment

- 1 On master node deviation of input data and layer 1 map is computed, and neurons for expansions are chosen (as described in Section 3.2). Important is that before start of learning all necessary preprocessing steps are done and input collection is ready in vector representation based on tfidf terms weighting scheme. Then from the input collection related vectors are selected (which are needed for particular expanded neurons) and distributed on working nodes. Using GridMiner methods current list of available working nodes is retrieved. Number of nodes is important parameter for distribution.
- 2 Distributed vectors are then used as inputs for GHSOM algorithm which runs on particular nodes in order to create hierarchical submodel. When end condition is reached (maximal depth or nothing to expand), particularly created GHSOM submodels are returned to the main node.
- 3 Returned parts of GHSOM model are merged (on the main node) into one complete model.

Every clustering task contains identifier of neuron within layer 1 map, list of (identifiers of) input vectors mapped on this neuron and parameters of GHSOM algorithm. Assignment of clustering tasks to Grid nodes is following:

Let h is number of neurons to be expanded and u is number of available working nodes on Grid, then

- 1 If $h \leq u$, into tasks queues of first h nodes exactly one clustering task per queue is assigned.
- 2 If $h > u$, in first iteration u clustering tasks are assigned to first u nodes, in next iteration rest of the tasks is assigned similarly while is needed.

After assigned tasks are distributed, particular submodels are created on separate nodes. When node finishes all tasks in queue, his work is finished. If all submodels are returned to the master node, merging of model finishes whole process. Reference to "parent" map node of level 1 is set correctly to main map created at start of the process as well as references to "children" are correctly set to maps created on particular nodes. Then final merged model is saved as persistent serialized Java object (important for next usage).

4.4 Combination of Local FCA Models Using Distributed Service-based Architecture

The implemented algorithm for distribution of FCA-based algorithm (presented in Section 3.3) on the Grid can be divided into two basic fragments - the server and client (worker) side. The method that we have designed, implemented and tested is sketched on Figure 3. The method works as follows.

In the first step, master node performs the following tasks: document pre-processing (tokenization, elimination of stop-words, stemming, term selection), document clustering (use of GHSOM algorithm), assigning each cluster (each map including clusters, neurons) to client. Second step is based on the client side (worker nodes) where each client will get particular clusters, and then local FCA model is created on each client followed by sending of local FCA hierarchies to master node. Last step is again performed on server side (master node), in this case server receives local FCA hierarchies from all clients and final merged FCA-based model is built.

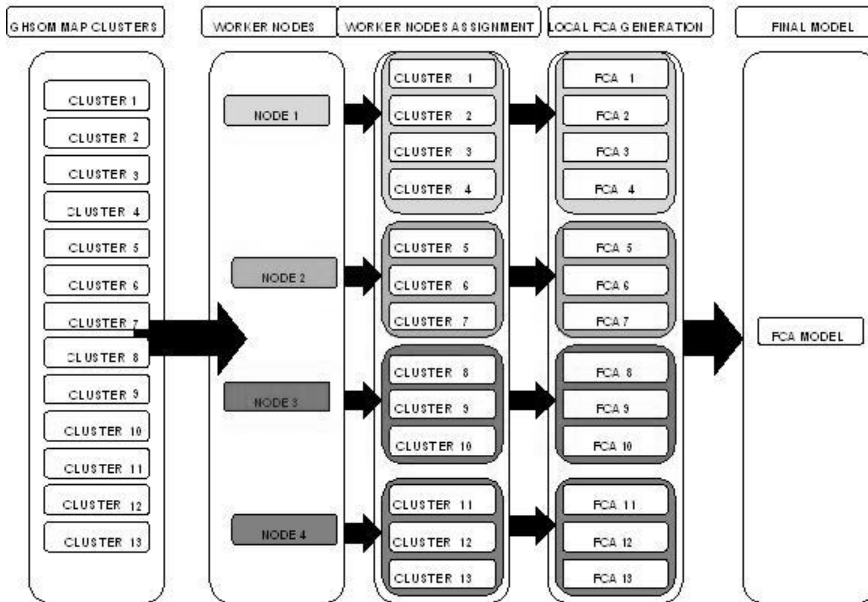


Figure 3

Distribution scheme for FCA-based problem decomposition method

5 Experiments

Two different data collections of text documents were used in this part of work. We used the collection “TIMES 60” which contains 420 articles from Times newspaper and Reuters ModApte dataset, which is standard corpus of textual documents that contains 12,902 documents. After pre-processing it contained 7769 documents and 2401 terms. We performed the experiments with the distributed algorithms on different workstations. In general, all of the workstations were Sun machines with different performance and different memory. Differences between types of text mining tasks will be emphasised.

5.1 Experiments with Decision Trees Induction

In this section, we present experiments performed on the local area network of the Institute of Scientific Computing in Vienna. As the experimental test bed, we used five workstations Sun Blade 1500, 1062 MHz Sparc CPU, 1.5 GB RAM connected by a 100 MBit network.

The main goal of the experiments was to prove, that the distribution of processes mentioned above, can reduce the time needed to construct the classification model. We started the experiments using the sequential version of the service, in order to compare the sequential version with the distributed one. The time to build the final classification model on a single machine using the ModApte dataset was measured three times and its mean value was 32.5 minutes. Then we performed the first series of the distributed service tests without using any optimization of distribution of categories to the worker nodes. According to the number of worker nodes, the master node assigned the equal number of categories to each worker node. The results show us the speedup of building the classification model using multiple nodes (see also Figure 4).

The detailed examination of the results and of the document collection proved that the time to build a complete classification model is significantly influenced by the working time of the first node. Examination of the dataset and workload of particular workers showed us that the first node always received a set of categories with the highest frequency of occurrences in the collection. It means that other worker nodes always finished the building of their partial models in a shorter time than the first one. It is caused by non-linear distribution of category occurrences in this collection. The most frequent category (category number 14) occurs in 2780 documents and it was always assigned to the first worker node. That was the reason, why the first worker node used much longer time to build-up the partial model.

After the first series of tests, we implemented the optimization of distribution of the categories to the worker nodes according to the frequency of category occurrences in the documents. Categories were sorted by this frequency and distributed to the worker nodes according to their frequency of occurrence, what means that each node was assigned with equal number of categories, but with a similar frequency of their occurrences. We run the same set of the experiments as in the first series and the results showed us more significant speedup using less worker nodes, see optimized bars in Figure 4. The best performance results were achieved using optimized distribution on 5 worker nodes (5.425 minutes), which was comparing to single machine computing time (32.5 minutes) almost 6 times faster. The minimal time to complete classification model is limited by the time of processing of the most frequent category - if this is assigned to a single worker node.

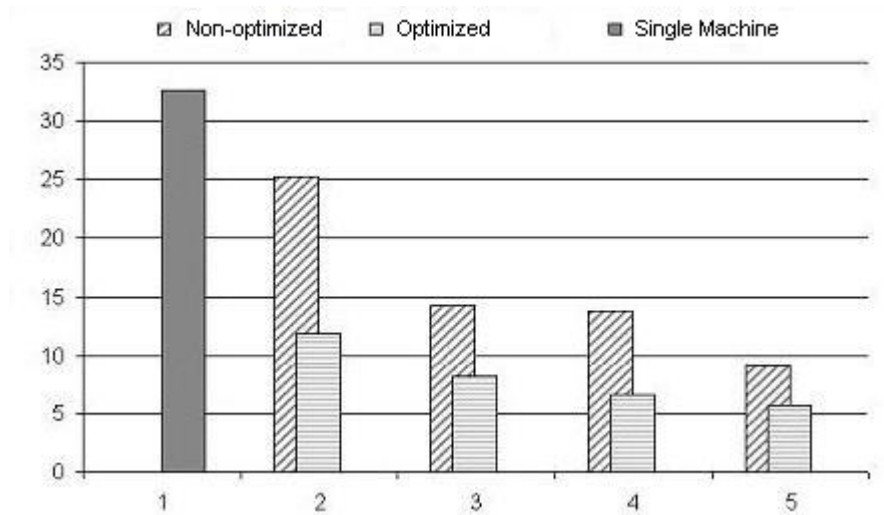


Figure 4

Experiments with distributed decision trees induction

5.2 Experiments with Distribution of GHSOM Maps Creation

The goal of experiments was in comparison of time complexity between sequential and distributed version of GHSOM algorithm. Experiments were realized in network of servers and workstations under usual working conditions (they were at the same time used also by other services). In order to get more precise results experiments are averaged from three identical tests. Number of computing nodes and parameter τ_m of GHSOM were changed during experiments.

Distributed version worked in testing Grid environment, which consisted of master server (4 x UltraSPARC-III 750 MHz, 8 GB RAM) and 6 SUN workstations, 100 Mbit/s network, data collections Times60 (420 documents) and Reuters-21578 (12 902 documents).

Experiments on Times collection were realized with τ_m 0.3, 0.6 and 0.8. First we started with sequential runs on one node. And then we tested distributed version for 0.3 τ_m with 2, 3, 4, 5 and 6 nodes (12 expanded neurons on layer 1 map). For parameter set to 0.6 and 0.8 only 2, 3 and 4 Grid nodes, because there were only 4 expanded neurons on layer 1 map. Graphical results of resulting computation times are shown on Figure 5.

For experiments with Reuters collection only τ_m with 0.6 and 0.8 was used and maximum number of nodes was 4. Results are shown on Figure 6.

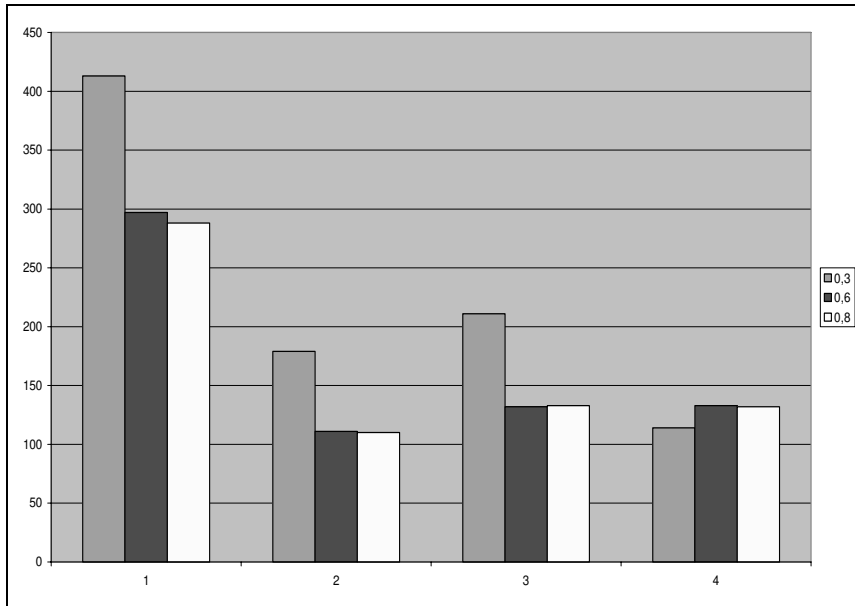


Figure 5

Graph of times (in seconds, y axis) for Times 60 collection for different number of nodes (x axis, max 4 nodes) with different values of τ_m parameter (legend – values 0.3, 0.6, 0.8)

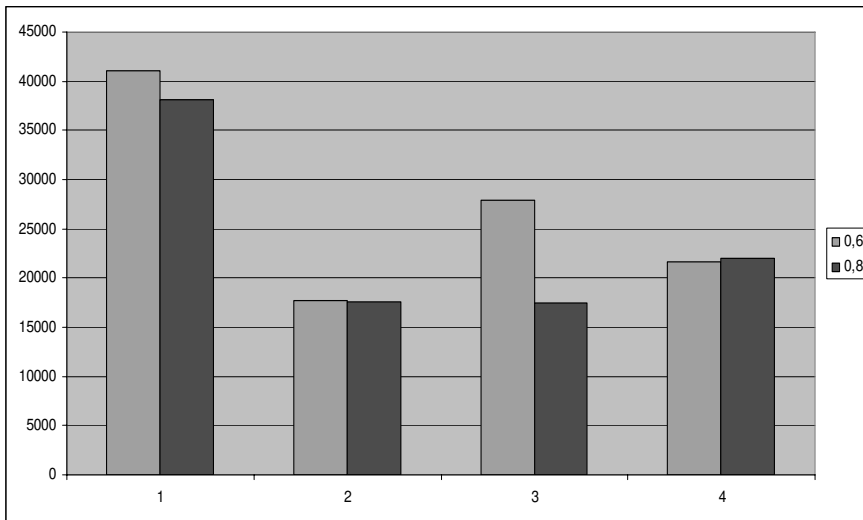


Figure 6

Graph of times (in seconds, y axis) for Reuters collection for different number of nodes (x axis, max 4 nodes) with different values of τ_m parameter (legend – values 0.6, 0.8)

The results of the experiments show interesting improvements in computation times of the algorithm in distributed version, but also the fact that addition of more worker nodes sometimes does not lead to better time reduction. The reasons could be unbalanced distribution of data for worker nodes, different values of variance error in learning from particular data parts and different computational power of Grid worker nodes. Better optimization of workers usage should be interesting for the next experiments. Critical point in such distributed version of GHSOM algorithm is creation of level 1 map (very often it is more then 50% of computing time). This means that further reductions of computation times are not possible with current distribution strategy. Combination of parallel building of first layer map (e.g. using computational cluster) and then distribution of this maps on the Grid could be helpful for another reduction of time complexity.

5.3 Experiments with FCA-based Distributed Approach

Again, both data collections of text documents were used in this part of work. Our main goal was to compare time consumption of the algorithm depending on the number of worker nodes. We used 9 different workstations deployed on the Grid. We performed various experiments with different values of *threshold* parameter 0.03, 0.05, 0.07 and 0.1. For each value of the threshold parameter we performed three runs of algorithm, final execution time of the algorithm was computed by averaging times of all three runs. The results of the experiments are depicted in the graphs.

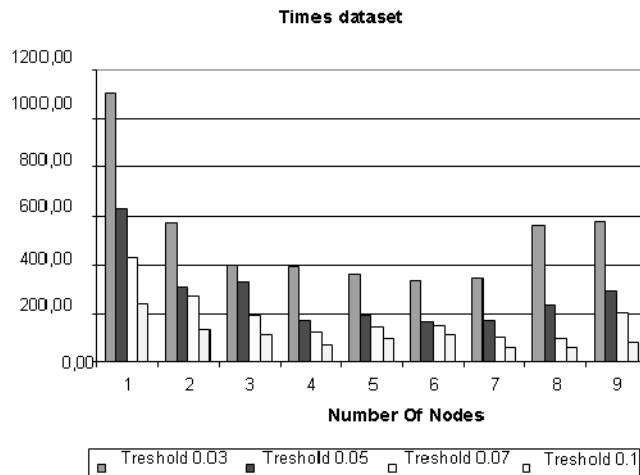


Figure 7
Experiment results on the Times dataset

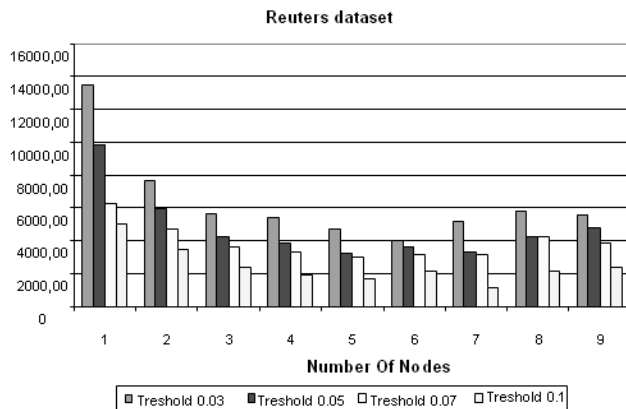


Figure 8

Experiment results on the Reuters dataset

We have compared number of the working nodes and its influence to the computing time of the merged FCA model generation. The results of experiments show that our distributed version has leads to clear reduction of computation time (even for small number of nodes). But the experiments in real environment also showed that increasing number of nodes leads (from some moment) to decreasing time complexity reduction. It is caused by heterogeneous computational power of involved worker nodes as well as their concurrent use by other, non-experimental tasks. More effective should be to optimize distribution of work according to the actual performance of particular nodes, i.e. dynamic distribution.

5.4 Discussion

Several questions could arise from the description of our grid-based approach. When discussing the difference between sequential and parallel running of tasks we have to emphasis that our approach cannot produce any information loss due to character of the decomposition and computing of the models. Text-mining results of the sequential and parallel run are therefore identical. Main aim of work presented in the paper is to provide the proof of concept for potential speedup of tasks computing that we have implemented. In this case proof of concept means that our approach is demonstrated on rather smaller number of computing units and applied on standard collections of text documents (Reuters) in order to prove that presented approach to distribution is scalable. Scalability of distributed algorithms is important issue, especially if applying them within large-scale distributed environment. Our experiments were aimed at algorithms behaviour in the testbed environment by increasing number of involved computing resources for which the selected datasets are sufficient. We assumed, that if our approach is proved to be scalable on our testbed, it can be applied in more large-scale fashion

(on larger datasets and using more computing resources) gaining similar results (speedup). Of course, bounds of scalability of these methods are data and environment-dependent. Our assumption based on experiments is that if data collections have approximately normal distribution of documents among the categories (in case of classification task), the scalability of our approach should be maintained also in larger scale. Similar assumptions can be expected also in case of clustering and FCA approaches. We have chosen three completely different text mining tasks, trying to cover text classification, text clustering and formal concept analysis tasks. Particular algorithms were chosen with respect to possibility of distributed implementation as our approach cannot be applied on several other algorithms.

Conclusions

Main aim of this article was to present the idea of suitable modifications and implementation of text mining services into the distributed Grid environment. Integration of the text mining services into the distributed service oriented system enables a plenty of various possibilities for building the distributed text mining scenarios. Using the Grid as a platform it is possible to access different distributed document collections and perform various text mining tasks. In this paper we focused on how to effectively use the Grid infrastructure by means of suitable decomposition of algorithms into the distributed fashion. We proposed three data-driven distributed methods for text mining: induction of the decision trees, GHSOM clustering algorithm and FCA method. One of the main goals of this work was to provide the proof of concept that proposed approach is well suited for distribution on the Grid, and results showed, that Grid-enabling of text mining process should considerably decrease time costs in comparison with sequential versions of these algorithms.

Acknowledgement

The work presented in the paper is supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic within the project No. 1/4074/07 "Methods for annotation, search, creation, and accessing knowledge employing metadata for semantic description of knowledge", and by the Slovak Research and Development Agency under the contracts No. RPEU-0011-06 (project PoZnaĽ) and No. APVV-0391-06 (project SEMCO-WS).

References

- [1] Foster I., Kesselman, C.: Computational Grids, The Grid – Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 1999
- [2] Luhn, H. P.: A Statistical Approach to Mechanized Encoding and Searching of Literary Information, in IBM Journal of Research and Development, 4:309-317, 1957
- [3] Quinlan, J. R.: Learning First-Order Definitions of Functions, in Journal of Artificial Intelligence Research, 5:139-161, 1996

-
- [4] Apte, C., Damerau, F., Weiss, S. M.: Towards Language Independent Automated Learning of Text Categorisation Models, in *Research and Development in Information Retrieval*, pp. 23-30, 1994
 - [5] Kohonen, T.: *Self-Organizing Maps*, Springer-Verlag, Berlin, 1995
 - [6] Dittenbach, M., Rauber, A., Merkl, D.: The Growing Hierarchical Self-Organizing Map, in *Proceedings of International Joint Conference on Neural Networks*, Como, Italy, 2000
 - [7] Ganter, B., Wille, R.: *Formal Concept Analysis*, Springer Verlag, 1997
 - [8] Krajci, S.: Clustering Algorithm Via Fuzzy Concepts, in *Proceedings of DATESO 2003 workshop*, Ostrava, Czech Republic, 2003, pp. 94-100
 - [9] Butka, P. Combination of Problem Reduction Techniques and Fuzzy FCA Approach for Building of Conceptual Models from Textual Documents (in Slovak), in *Znalosti 2006, 5th annual conference*, Ostrava, Czech Republic, 2006, pp. 71-82
 - [10] Belohlavek, R.: Concept Lattices and Formal Concept Analysis (in Czech), in *Znalosti 2004, 3rd annual conference*, Brno, Czech Rep., 2004, pp. 66-84
 - [11] Quan, T. T., Hui, S. C., Cao, T. H.: A Fuzzy FCA-based Approach to Conceptual Clustering for Automatic Generation of Concept Hierarchy on Uncertainty Data, in *Proceedings of CLA conference*, Ostrava, Czech Republic, 2004, pp. 1-12
 - [12] Bednar, P., Butka, P., Paralic., J.: Java Library for Support of Text Mining and Retrieval, in *Proceedings of Znalosti 2005, 4th annual conference*, Stara Lesna, Slovakia, 2005, pp. 162-169
 - [13] Brezany, P., Janciak, I., Woehrer, A., Tjoa, A. M.: Gridminer: A Framework for Knowledge Discovery on the Grid - From a Vision to Design and Implementation, in *Cracow Grid Workshop*, Cracow, Poland, 2004
 - [14] Brezany, P., Janciak, I., Sarnovsky, M.: Text Mining within the GridMiner Framework, in *2nd Dialogue Workshop*, Edinburgh, GB, 2006
 - [15] Janciak, I., Sarnovsky, M., Tjoa, A. M., Brezany, P.: Distributed Classification of Textual Documents on the Grid, in *High Performance Computing and Communications, HPCC 2006, LNCS 4208*, Munich, Germany, September 13-15, 2006, pp. 710-718
 - [16] Sarnovský, M., Butka, P., Safko, V.: Distributed Clustering of Textual Documents in the Grid Environment (in Slovak), in *Znalosti 2008, 7th annual conference*, Bratislava, Slovakia, 2008, pp. 192-203
 - [17] Butka, P., Sarnovský, M., Bednár, P. One Approach to Combination of FCA-based Local Conceptual Models for Text Analysis - Grid-based

Approach, in Proceedings of SAMI 2008, IEEE conference, Herlany, Slovakia, 2008, pp. 131-135

- [18] Butka, P., Zeher, M. Simple Approach to Combination of FCA-based Local Conceptual Models for Text Analysis. In Proceedings of the 7th International Workshop on Data Analysis, WDA 2006, Košice, Slovakia, 2006, pp. 1-10

Adaptive Fuzzy Control Design

Martin Kratmüller

SIEMENS PSE sro Slovakia
Dúbravská cesta 4, 845 37 Bratislava, Slovak Republic
E-mail: martin.kratmueller@siemens.com

Abstract: An application of fuzzy systems to nonlinear system adaptive control design is proposed in this paper. The fuzzy system is constructed to approximate the nonlinear system dynamics. Based on this fuzzy approximation suitable adaptive control laws and appropriate parameter update algorithms for nonlinear uncertain (or unknown) systems are developed to achieve H_∞ tracking performance. It is shown that the effects of approximation errors and external disturbance can be attenuated to a specific attenuation level using the proposed adaptive fuzzy control scheme. The nonlinear gradient law guarantees the convergence of the training algorithm.

Keywords: adaptive fuzzy control, Riccati equation, uncertain system, nonlinear systems

1 Introduction

Fuzzy logic controllers are in general considered being applicable to plants that are mathematically poorly understood and where the experienced human operators are available [1]. In indirect adaptive fuzzy control, the fuzzy logic systems are used to model the plant. Then a controller is constructed assuming that the fuzzy logic system approximately represents the true plant.

Feedback linearization techniques for nonlinear control system design have been developed in the last two decades [2], [3]. However, these techniques can only be applied to nonlinear systems whose parameters are known exactly. If the nonlinear system contains unknown or uncertain parameters then the feedback linearization is no longer utilisable. In this situation, the adaptive strategies are used to simplify the problem and to allow a suitable solution. At present, a number of adaptive control design techniques for nonlinear systems based on the feedback linearization can be found in literature [4], [5]. These approaches simplify the nonlinear systems by assuming either linearly or nonlinearly parametrized structures. However, these assumptions are not sufficient for many practical applications. Recently, the fuzzy systems have been employed successfully in the adaptive control design problems of nonlinear systems. According to the universal

approximation theorem [6], [7], many important adaptive fuzzy-based control schemes have been developed to incorporate the expert information directly and systematically and various stable performance criteria are guaranteed by theoretical analysis [6], [8]-[12].

In this paper we combine the characteristics of fuzzy systems, the technique of feedback linearization, the adaptive control scheme and the H_∞ optimal control theory with aim to solve the tracking control design problem for nonlinear systems with bounded unknown or uncertain parameters and external disturbances. H_∞ optimal control theory is well known as an efficient tool for robust stabilization and disturbance rejection problems [13], [14].

More specifically, we propose the fuzzy adaptive algorithm equipped with a gradient projection law. The resulting controller performances can be improved by incorporating some linguistic rules describing the plant dynamic behavior.

The paper is organized as follows. First, the problem formulation is presented in Section 2. In Section 3, the adaptive fuzzy control is proposed. Simulation results for the proposed control concept are shown in Section 4. Finally, the paper is concluded in Section 5.

2 Problem Statement

We consider the n -th order nonlinear dynamic single input single output (SISO) system with $n \geq 2$ of the following form

$$\begin{aligned} \dot{x}_1 &= x_2 \\ &\vdots \\ \dot{x}_n &= f(\underline{x}) + g(\underline{x})u + d \\ y &= x_1 \end{aligned} \tag{1}$$

or equivalently

$$\begin{aligned} \dot{\mathbf{x}}^{(n)} &= f(\mathbf{x}, \dot{\mathbf{x}}, \dots, \mathbf{x}^{(n-1)}) + g(\mathbf{x}, \dot{\mathbf{x}}, \dots, \mathbf{x}^{(n-1)})u + d \\ y &= \mathbf{x} \end{aligned} \tag{2}$$

where $\underline{x} = [x_1, x_2, \dots, x_n]^T$ represents the state vector, u is the control input, y and d denote the system output and the external disturbance, respectively. All elements of the state vector \underline{x} are assumed to be available and the external

disturbance d is assumed to be bounded but unknown or uncertain. At the beginning $f(\underline{x})$ and $g(\underline{x})$ are assumed to be smooth and $g(\underline{x}) \neq 0$ for \underline{x} in certain controllability region $U_c \subset \mathbb{R}^n$. Without loss of generality we suppose that $g(\underline{x}) > 0$, but the analysis throughout this paper can be easily tailored to systems with $g(\underline{x}) < 0$. Differentiating the output y with respect to time for n times gives the following input/output form

$$y^{(n)} = f(\underline{x}) + g(\underline{x})u + d \quad (3)$$

Note that the above system has a relative degree of n .

Remark 1. For more general nonlinear system

$$\begin{aligned} \dot{\underline{z}} &= F(\underline{z}) + G(\underline{z})u + d' \\ y &= H(\underline{z}) \end{aligned} \quad (4)$$

where $\underline{z} \in \mathbb{R}^n$, $u, v \in \mathbb{R}$, $F(\underline{z})$, $G(\underline{z})$ and $H(\underline{z})$ are smooth functions, we say that the system has a relative degree of m if m is the smallest integer such that $L_G L_F^{m-1} H \neq 0$.

We obtain [2]

$$\begin{aligned} y^{(m)} &= L_F^m H + L_G L_F^{m-1} H u + L_{F+Gu+d}^{m-1} L_d H \\ &\quad + \sum_{k=1}^{m-1} L_{F+Gu+d}^{k-1} L_d L_F^{m-k} H \end{aligned} \quad (5)$$

where $L_F(\cdot)$, and $L_G(\cdot)$ denote the Lie derivatives with respect to F and G , respectively. If we let $y = x_1$, then (5) can be rewritten as the input/output form of (3).

If $f(\underline{x})$ and $g(\underline{x})$ are known, a nonlinear tracking control can be obtained. Let y_r be the desired continuous differentiable uniformly bounded trajectory and let

$$e = y - y_r \quad (6)$$

be the tracking error. Then employing the technique of feedback linearization [2] the following suitable control law can be derived to achieve the tracking control goal

$$\mathbf{u} = \frac{1}{g(\underline{x})} \left[-f(\underline{x}) + \mathbf{u}_a + \mathbf{v} \right] \quad (7)$$

where \mathbf{u}_a is an auxiliary control variable [13, optimal control] yet to be specified and

$$\mathbf{v} = y_r^{(n)} + k_1 \left(y_r^{(n-1)} - y^{(n-1)} \right) + \dots + k_n (y_r - y) \quad (8)$$

Note that the coefficients k_1, \dots, k_n are positive constants to be assigned such that the polynomial $s^n + k_1 s^{n-1} + \dots + k_n$ is Hurwitz. As a result, the system error dynamic has the following input/output form

$$e^{(n)} + k_1 e^{(n-1)} + \dots + k_n e = \mathbf{u}_a + \mathbf{d} \quad (9)$$

which can be represented in space form as

$$\dot{\underline{e}} = \mathbf{\Lambda}_c \underline{e} + \underline{b}_c (\mathbf{u}_a + \mathbf{d}) \quad (10)$$

where

$$\mathbf{\Lambda}_c = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -k_n & -k_{n-1} & & \dots & & -k_1 \end{bmatrix} \quad (11)$$

$$\underline{b}_c = [0 \quad \dots \quad 0 \quad 1]^T \quad (12)$$

$$\underline{e} = [e \quad \dots \quad e^{(n-2)} \quad e^{(n-1)}]^T \quad (13)$$

The above mentioned design method is useful only if $f(\underline{x})$ and $g(\underline{x})$ are known exactly. If $f(\underline{x})$ and $g(\underline{x})$ are unknown, then adaptive strategies must be employed. Let us now discuss a fuzzy system based adaptive algorithm.

First, we employ two fuzzy systems $\hat{f}(\underline{x}|\underline{\theta}_f)$ and $\hat{g}(\underline{x}|\underline{\theta}_g)$ [15] to approximate (or model) the nonlinear functions $f(\underline{x})$ and $g(\underline{x})$ of the system (1).

In this article is used the set of fuzzy systems with singleton fuzzifier, product inference, centroid defuzzifier, triangular antecedent membership function and singleton consequent membership function with n inputs of $x_i \in [c_{x_i} - k_{x_i}, c_{x_i} + k_{x_i}]$ for $i = 1, \dots, n$ and $\bar{u} \in [0, 1]$ as the normalized output. The generalized expression of the class of the fuzzy controllers can be written as

$$\bar{u} = \sum_{i_1=1}^2 \cdots \sum_{i_n=1}^2 N_{i_1 \cdots i_n} x_1^{i_1-1} \cdots x_n^{i_n-1} \quad (14)$$

$$N_{i_1 \cdots i_n} = \frac{\left[\sum_{j_1=1}^2 \cdots \sum_{j_n=1}^2 R_{j_1 \cdots j_n} K_{j_1 \cdots j_n} C_{j_1 \cdots j_n} \right]}{2^n \prod_{i=1}^n k_{x_i}} \quad (15)$$

$$C_{j_1 \cdots j_n} = \left[\frac{(-1)^{j_1}}{k_{x_1} - (-1)^{j_1} c_{x_1}} \right]^{i_1-1} \cdots \left[\frac{(-1)^{j_n}}{k_{x_n} - (-1)^{j_n} c_{x_n}} \right]^{i_n-1} \quad (16)$$

$$K_{j_1 \cdots j_n} = [k_{x_1} - (-1)^{j_1} c_{x_1}] \cdots [k_{x_n} - (-1)^{j_n} c_{x_n}] \quad (17)$$

On the other hand given the coefficients of the explicit form $N_{i_1 \cdots i_n}$ we can reconstruct the rule base from the generalized expression of the class of fuzzy systems [16] by using the following theorem.

Theorem 1: For a class of FLS with singleton fuzzifier, product inference, centroid defuzzifier, triangular antecedent membership function and singleton consequent membership function, i.e. given the coefficients of the explicit form, i.e. $N_{i_1 \cdots i_n}$, the control function can be expressed in terms of fuzzy rules as

$$R_{j_1 \cdots j_n} = \sum_{i_1=1}^2 \cdots \sum_{i_n=1}^2 N_{i_1 \cdots i_n} D_{j_1 \cdots j_n} \quad (18)$$

with

$$D_{j_1 \cdots j_n} = [c_{x_1} + (-1)^{j_1} k_{x_1}]^{i_1-1} \cdots [c_{x_n} + (-1)^{j_n} k_{x_n}]^{i_n-1} \quad (19)$$

Proof: The proof is found by directly expanding terms and comparing coefficients. For details, please refer [16].

Therefore, one can express an equation in the form of generalized multilinear equations, such as polynomials, exactly as a rule base of FLS. Theorem 1 is useful

in cases where the implementation of an FLS performs inference on a given fuzzy rule base but without any numerical computation capability.

Now, we can express the fuzzy controller in the form of fuzzy IF-THEN rules.

1) For the nonlinear-cancellation fuzzy controller of $f(\underline{x})$

RULE i: IF x_1 is $A_1^{x_1}$ and ... and x_n is $A_n^{x_n}$, THEN $\bar{u}_f = R_i^f$

2) For the nonlinear-cancellation fuzzy controller of $g(\underline{x})$

RULE i: IF x_1 is $A_1^{x_1}$ and ... and x_n is $A_n^{x_n}$, THEN $\bar{u}_g = R_i^g$

The generalized expression of the class of the fuzzy approximators for nonlinear term cancelation with input x can be written as controller for pole-placement

$$\bar{u}_f = \sum_{i_1=1}^2 \cdots \sum_{i_n=1}^2 N_{i_1 \cdots i_n}^f x_1^{i_1-1} \cdots x_n^{i_n-1} \quad (20)$$

$$\bar{u}_g = \sum_{i_1=1}^2 \cdots \sum_{i_n=1}^2 N_{i_1 \cdots i_n}^g x_1^{i_1-1} \cdots x_n^{i_n-1} \quad (21)$$

So terms for $\hat{f}(\underline{x} | \underline{\theta}_f)$ and $\hat{g}(\underline{x} | \underline{\theta}_g)$ can be written as

$$\hat{f}(\underline{x} | \underline{\theta}_f) = \underline{\theta}_f^T \underline{\omega}_x \quad (22)$$

$$\text{with } \underline{\theta}_f^T = (\underline{k}_{f_b}^T, \underline{k}_{f_c}^T)$$

$$\text{and } \underline{\omega}_x^T = (\underline{x}^T, \underline{x}_c^T)$$

and

$$\hat{g}(\underline{x} | \underline{\theta}_g) = \underline{\theta}_g^T \underline{\omega}_x \quad (23)$$

$$\text{with } \underline{\theta}_g^T = (\underline{k}_{g_b}^T, \underline{k}_{g_c}^T)$$

$$\text{and } \underline{\omega}_x^T = (\underline{x}^T, \underline{x}_c^T)$$

$$\text{with } \underline{k}_{f_b}^T = [k_1^f, \cdots, k_n^f] \text{ and } \underline{k}_{g_b}^T = [k_1^g, \cdots, k_n^g]$$

where

$$\begin{aligned}
\mathbf{k}_1^f &= 2\mathbf{N}_{211\dots111}^f & \mathbf{k}_1^g &= 2\mathbf{N}_{211\dots111}^g \\
\mathbf{k}_2^f &= 2\mathbf{N}_{121\dots111}^f & \mathbf{k}_2^g &= 2\mathbf{N}_{121\dots111}^g \\
\mathbf{k}_{n-1}^f &= 2\mathbf{N}_{111\dots121}^f & \mathbf{k}_{n-1}^g &= 2\mathbf{N}_{111\dots121}^g \\
\mathbf{k}_n^f &= 2\mathbf{N}_{111\dots112}^f & \mathbf{k}_n^g &= 2\mathbf{N}_{111\dots112}^g
\end{aligned}$$

The composite state vector $\underline{\mathbf{x}}_c$ and the associated parameter vectors $\underline{\mathbf{k}}_{f_c}$, $\underline{\mathbf{k}}_{g_c}$ are defined as

$$\underline{\mathbf{x}}_c^T = (\mathbf{r}\mathbf{x}_1\mathbf{x}_2 \dots \mathbf{x}_n, \mathbf{r}\mathbf{x}_1\mathbf{x}_2 \dots \mathbf{x}_{n-1}, \dots, \mathbf{x}_{n-1}\mathbf{x}_n, 1) \quad (24)$$

$$\underline{\mathbf{k}}_{f_c}^T = (\mathbf{k}_{n+1}^f, \mathbf{k}_{n+2}^f, \dots, \mathbf{k}_{n+n_c-1}^f, \mathbf{k}_{n+n_c}^f) \quad (25)$$

$$\underline{\mathbf{k}}_{g_c}^T = (\mathbf{k}_{n+1}^g, \mathbf{k}_{n+2}^g, \dots, \mathbf{k}_{n+n_c-1}^g, \mathbf{k}_{n+n_c}^g) \quad (26)$$

where

$$\begin{aligned}
\mathbf{k}_{n+1}^f &= 2\mathbf{N}_{222\dots222}^f & \mathbf{k}_{n+1}^g &= 2\mathbf{N}_{222\dots222}^g \\
\mathbf{k}_{n+2}^f &= 2\mathbf{N}_{222\dots221}^f & \mathbf{k}_{n+2}^g &= 2\mathbf{N}_{222\dots221}^g \\
\mathbf{k}_{n+n_c-1}^f &= 2\mathbf{N}_{111\dots122}^f & \mathbf{k}_{n+n_c-1}^g &= 2\mathbf{N}_{111\dots122}^g \\
\mathbf{k}_{n+n_c}^f &= 2\mathbf{N}_{111\dots111}^f & \mathbf{k}_{n+n_c}^g &= 2\mathbf{N}_{111\dots111}^g
\end{aligned}$$

with $n_c = 2^{n+1} - (n+1)$.

Let

$$\underline{\theta}_f^* = \arg \min_{\underline{\theta}_f} \max_{\underline{\mathbf{x}}} |\hat{\mathbf{f}}(\underline{\mathbf{x}}, \underline{\theta}_f)| \quad (27)$$

$$\underline{\theta}_g^* = \arg \min_{\underline{\theta}_g} \max_{\underline{\mathbf{x}}} |\hat{\mathbf{g}}(\underline{\mathbf{x}}, \underline{\theta}_g)| \quad (28)$$

be the best parameter approximation of $\underline{\theta}_f$ and $\underline{\theta}_g$, respectively, and let

$$\underline{\phi}_f = \underline{\theta}_f - \underline{\theta}_f^*, \quad \underline{\phi}_g = \underline{\theta}_g - \underline{\theta}_g^* \quad (29)$$

be the corresponding parameter estimation errors. Then using the certainty equivalence principle [5] the following fuzzy adaptive control law is derived

$$\mathbf{u} = \frac{1}{\widehat{\mathbf{g}}(\underline{\mathbf{x}}, \underline{\boldsymbol{\theta}}_g)} \left[-\widehat{\mathbf{f}}(\underline{\mathbf{x}}, \underline{\boldsymbol{\theta}}_f) + \mathbf{u}_a + \mathbf{v} \right] \quad (30)$$

Applying this control law to the system (1) yields

$$\begin{aligned} \mathbf{y}^{(n)} &= \mathbf{f}(\underline{\mathbf{x}}) + \mathbf{g}(\underline{\mathbf{x}})\mathbf{u} + \mathbf{d} \\ &= \mathbf{f}(\underline{\mathbf{x}}) + \mathbf{g}(\underline{\mathbf{x}})\mathbf{u} - \widehat{\mathbf{g}}(\underline{\mathbf{x}}, \underline{\boldsymbol{\theta}}_g)\mathbf{u} + \widehat{\mathbf{g}}(\underline{\mathbf{x}}, \underline{\boldsymbol{\theta}}_g)\mathbf{u} + \mathbf{d} \\ &= \left(\mathbf{f}(\underline{\mathbf{x}}) - \widehat{\mathbf{f}}(\underline{\mathbf{x}}, \underline{\boldsymbol{\theta}}_f) \right) + \left(\mathbf{g}(\underline{\mathbf{x}}) - \widehat{\mathbf{g}}(\underline{\mathbf{x}}, \underline{\boldsymbol{\theta}}_g) \right) \mathbf{u} + \mathbf{u}_a + \mathbf{v} + \mathbf{d} \end{aligned} \quad (31)$$

By means of the best approximation (using the universal approximation theorem [6], [7], [17]), the above equation can be rewritten as

$$\begin{aligned} \dot{\underline{\mathbf{e}}} &= \mathbf{\Lambda}_c \underline{\mathbf{e}} + \mathbf{b}_c \left[\left(\widehat{\mathbf{f}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_f) - \widehat{\mathbf{f}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_f^*) \right) \right. \\ &\quad \left. + \left(\widehat{\mathbf{g}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_g) - \widehat{\mathbf{g}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_g^*) \right) \mathbf{u} \right] + \mathbf{b}_c [\mathbf{u}_a + \mathbf{w}] \end{aligned} \quad (32)$$

where

$$\mathbf{w} = \left(\mathbf{f}(\underline{\mathbf{x}}) - \widehat{\mathbf{f}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_f^*) \right) + \left(\mathbf{g}(\underline{\mathbf{x}}) - \widehat{\mathbf{g}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_g^*) \right) \mathbf{u} + \mathbf{d} \quad (33)$$

In order to track the desired signal \mathbf{y}_r , the fuzzy systems $\widehat{\mathbf{f}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_f)$ and $\widehat{\mathbf{g}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_g)$ should be trained to achieve $\widehat{\mathbf{f}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_f^*)$ and $\widehat{\mathbf{g}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_g^*)$ respectively, so that the term

$$\left[\left(\widehat{\mathbf{f}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_f^*) - \widehat{\mathbf{f}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_f) \right) + \left(\widehat{\mathbf{g}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_g^*) - \widehat{\mathbf{g}}(\underline{\mathbf{x}} | \underline{\boldsymbol{\theta}}_g) \right) \mathbf{u} \right] = 0 \quad (34)$$

The effect of \mathbf{w} , denoting the sum of the approximation errors and external disturbances in the above error dynamics equation, is crucial and will be attenuated by \mathbf{u}_a . Fortunately, the \mathbf{H}_∞ control design approach [12] can be efficiently employed to attenuate the effect of \mathbf{w} in the error dynamic system (32). Our solution utilizes the concept of \mathbf{H}_∞ tracking performance to deal with the robust adaptive tracking control problem. Then, the problem we are investigating becomes that of finding an adaptive scheme for \mathbf{u}_a , $\underline{\boldsymbol{\theta}}_f$ and $\underline{\boldsymbol{\theta}}_g$ to achieve the following \mathbf{H}_∞ tracking performance [12], [18]

$$\int_0^T \underline{e}^T \mathbf{Q} \underline{e} dt \leq \underline{e}^T(0) \mathbf{P} \underline{e}(0) + \frac{1}{\gamma_f} \underline{\phi}_f^T(0) \underline{\phi}_f(0) + \frac{1}{\gamma_g} \underline{\phi}_g^T(0) \underline{\phi}_g(0) + \rho^2 \int_0^T \mathbf{w}^T \mathbf{w} dt \quad (35)$$

for appropriate positive definite weighting matrices $\mathbf{Q} = \mathbf{Q}^T$, $\mathbf{P} = \mathbf{P}^T$, positive weighting factors γ_f and γ_g , prescribed attenuation level ρ and time T . In the inequality (35), T is the terminal time of the control effort and can take any finite or infinite value. The initial errors $\underline{e}(0)$, $\underline{\phi}_f(0)$ and $\underline{\phi}_g(0)$ are considered to be free of the disturbances which can influence the tracking error \underline{e} . The physical meaning of (35) is that the effect of w on the tracking error \underline{e} is attenuated by a factor ρ from an energy point of view. In general ρ is a small value less than 1.

Remark 2. From the above analysis, we note the following

- In the case of $\rho \rightarrow \infty$, (35) becomes the H_2 tracking performance without consideration of disturbance attenuation [12].
- The weighting factors γ_f and γ_g are called the adaptive gains of $\underline{\theta}_f$ and $\underline{\theta}_g$ update algorithms, respectively. It can be seen from (35), that the larger the value of γ_f , the smaller the effect of $\underline{\phi}_f(0)$ on the tracking error \underline{e} . Similar argument for $\underline{\phi}_g(0)$ can also be made. However, it is easy to see that large values of γ_f or γ_g will cause $\underline{\theta}_f$ and $\underline{\theta}_g$ to change rapidly. This may be harmful to the system.

3 Adaptive Fuzzy Control

The following theorem gives the solution of the adaptive H_∞ tracking problem for the SISO nonlinear system (1).

Theorem 2. Consider the nonlinear system (1) with unknown or uncertain $f(\underline{x})$ and $g(\underline{x})$. If the following adaptive fuzzy control law is adopted

$$\mathbf{u} = \frac{1}{\hat{g}(\underline{x}|\underline{\theta}_g)} \left[-\hat{f}(\underline{x}|\underline{\theta}_f) + \mathbf{u}_a + \mathbf{v} \right] \quad (36)$$

with

$$\underline{u}_a = -\frac{1}{r} \underline{b}_c^T \underline{P} \underline{e} \quad (37)$$

$$\dot{\underline{\theta}}_f = \gamma_f \underline{e}^T \underline{P} \underline{b}_c \underline{\omega}_x \quad (38)$$

$$\dot{\underline{\theta}}_g = \gamma_g \underline{e}^T \underline{P} \underline{b}_c \underline{\omega}_x \underline{u} \quad (39)$$

where the signal \underline{v} is given by (8), r is a positive scalar, the fuzzy systems $\hat{f}(\underline{x}|\underline{\theta}_f)$ and $\hat{g}(\underline{x}|\underline{\theta}_g)$ are defined by (22), (23) and the positive definite matrix $\underline{P} = \underline{P}^T$ is the solution of the Riccati-like equation

$$\underline{\Lambda}_c \underline{P}^T + \underline{P} \underline{\Lambda}_c + \underline{Q} - \frac{2}{r} \underline{P} \underline{b}_c \underline{b}_c^T \underline{P} + \frac{1}{\rho^2} \underline{P} \underline{b}_c \underline{b}_c^T \underline{P} = 0 \quad (40)$$

then the H_∞ tracking performance in (35) is achieved for a prescribed attenuation level ρ .

Proof. Consider the Lyapunov function in the form

$$\underline{V} = \frac{1}{2} \underline{e}^T \underline{P} \underline{e} + \frac{1}{2\gamma_f} \underline{\phi}_f^T \underline{\phi}_f + \frac{1}{2\gamma_g} \underline{\phi}_g^T \underline{\phi}_g \quad (41)$$

Taking the time derivative of \underline{V} along the trajectory of the error dynamic (8), we have

$$\begin{aligned} \dot{\underline{V}} &= \frac{1}{2} \dot{\underline{e}}^T \underline{P} \underline{e} + \frac{1}{2} \underline{e}^T \underline{P} \dot{\underline{e}} + \frac{1}{2\gamma_f} \dot{\underline{\phi}}_f^T \underline{\phi}_f + \frac{1}{2\gamma_f} \underline{\phi}_f^T \dot{\underline{\phi}}_f \\ &\quad + \frac{1}{2\gamma_g} \dot{\underline{\phi}}_g^T \underline{\phi}_g + \frac{1}{2\gamma_g} \underline{\phi}_g^T \dot{\underline{\phi}}_g \\ &= \frac{1}{2} \underline{e}^T \underline{\Lambda}_c^T \underline{P} \underline{e} + \frac{1}{2} \underline{e}^T \underline{P} \underline{\Lambda}_c \underline{e} + \frac{1}{2} \underline{e}^T \underline{P} \underline{b}_c \underline{u}_a + \frac{1}{2} \underline{u}_a^T \underline{b}_c^T \underline{P} \underline{e} \\ &\quad + \frac{1}{2} \left[\left(\hat{f}(\underline{x}|\underline{\theta}_f^*) - \hat{f}(\underline{x}|\underline{\theta}_f) \right)^T + \left(\hat{g}(\underline{x}|\underline{\theta}_g^*) - \hat{g}(\underline{x}|\underline{\theta}_g) \right)^T \underline{u} \right] \underline{b}_c^T \underline{P} \underline{e} \\ &\quad + \frac{1}{2} \underline{e}^T \underline{P} \underline{b}_c \left[\left(\hat{f}(\underline{x}|\underline{\theta}_f^*) - \hat{f}(\underline{x}|\underline{\theta}_f) \right) + \left(\hat{g}(\underline{x}|\underline{\theta}_g^*) - \hat{g}(\underline{x}|\underline{\theta}_g) \right) \underline{u} \right] \\ &\quad + \frac{1}{2} \underline{e}^T \underline{P} \underline{b}_c \underline{w} + \frac{1}{2} \underline{w}^T \underline{b}_c^T \underline{P} \underline{e} + \frac{1}{2\gamma_f} \dot{\underline{\phi}}_f^T \underline{\phi}_f + \frac{1}{2\gamma_f} \underline{\phi}_f^T \dot{\underline{\phi}}_f \\ &\quad + \frac{1}{2\gamma_g} \dot{\underline{\phi}}_g^T \underline{\phi}_g + \frac{1}{2\gamma_g} \underline{\phi}_g^T \dot{\underline{\phi}}_g \end{aligned} \quad (42)$$

Using (22), (23), (28), (39) and the fact that

$$\dot{\underline{\phi}}_f = \dot{\underline{\theta}}_f, \quad \dot{\underline{\phi}}_g = \dot{\underline{\theta}}_g \quad (43)$$

we obtain

$$\begin{aligned} \dot{V} = & \frac{1}{2} \underline{e}^T \left[\underline{\Lambda}_c^T \mathbf{P} + \mathbf{P} \underline{\Lambda}_c - \frac{2}{r} \mathbf{P} \underline{b}_c \underline{b}_c^T \mathbf{P} \right] \underline{e} \\ & - \left[\underline{\phi}_f^T \underline{\omega}_x + \mathbf{u} \underline{\phi}_g^T \underline{\omega}_x \right] \underline{b}_c^T \mathbf{P} \underline{e} \\ & + \frac{1}{\gamma_f} \underline{\phi}_f^T \dot{\underline{\phi}}_f + \frac{1}{\gamma_g} \underline{\phi}_g^T \dot{\underline{\phi}}_g + \frac{1}{2} \mathbf{w}^T \underline{b}_c^T \mathbf{P} \underline{e} + \frac{1}{2} \underline{e}^T \mathbf{P} \underline{b}_c \mathbf{w} \end{aligned} \quad (44)$$

Introducing (40) into (44) implies

$$\begin{aligned} \dot{V} = & -\frac{1}{2} \underline{e}^T \mathbf{Q} \underline{e} - \frac{1}{2\rho^2} \underline{e}^T \mathbf{P} \underline{b}_c \underline{b}_c^T \mathbf{P} \underline{e} \\ & - \underline{\phi}_f^T \left(\underline{\omega}_x \underline{b}_c^T \mathbf{P} \underline{e} - \frac{1}{\gamma_f} \dot{\underline{\theta}}_f \right) \\ & - \underline{\phi}_g^T \left(\mathbf{u} \underline{\omega}_x \underline{b}_c^T \mathbf{P} \underline{e} - \frac{1}{\gamma_g} \dot{\underline{\theta}}_g \right) \\ & + \frac{1}{2} \mathbf{w}^T \underline{b}_c^T \mathbf{P} \underline{e} + \frac{1}{2} \underline{e}^T \mathbf{P} \underline{b}_c \mathbf{w} \end{aligned} \quad (45)$$

Using the adaptation laws (38) and (39), equation (45) can be rewritten into the form

$$\begin{aligned} \dot{V} = & -\frac{1}{2} \left(\frac{1}{\rho} \underline{b}_c^T \mathbf{P} \underline{e} - \rho \mathbf{w} \right)^T \left(\frac{1}{\rho} \underline{b}_c^T \mathbf{P} \underline{e} - \rho \mathbf{w} \right) + \frac{1}{2} \rho^2 \mathbf{w}^T \mathbf{w} \\ & - \frac{1}{2} \underline{e}^T \mathbf{Q} \underline{e} \\ \leq & -\frac{1}{2} \underline{e}^T \mathbf{Q} \underline{e} + \frac{1}{2} \rho^2 \mathbf{w}^T \mathbf{w} \end{aligned} \quad (46)$$

Integrating the above equation from 0 to T yields

$$V(T) - V(0) \leq -\frac{1}{2} \int_0^T \underline{e}^T \mathbf{Q} \underline{e} dt + \frac{1}{2} \rho^2 \int_0^T \mathbf{w}^T \mathbf{w} dt \quad (47)$$

Since $V(T) \geq 0$ inequality (47) implies that

$$\int_0^T \underline{e}^T \underline{Q} \underline{e} dt \leq \underline{e}^T(0) \underline{P} \underline{e}(0) + \frac{1}{\gamma_f} \underline{\phi}_f^T(0) \underline{\phi}_f(0) + \frac{1}{\gamma_g} \underline{\phi}_g^T(0) \underline{\phi}_g(0) + \rho^2 \int_0^T w^T w dt \quad (48)$$

This is the H_∞ tracking performance of (35).

Q.E.D.

Remark 3. If w is bounded, then the H_∞ tracking performance will be improved as the prescribed attenuation level ρ is decreased.

Remark 4. The Riccati-like equation (40) can be rewritten into the form

$$\underline{P} \underline{\Lambda}_c + \underline{\Lambda}_c^T \underline{P} + \underline{P} \underline{b}_c \left(\frac{1}{\rho^2} - \frac{2}{r} \right) \underline{b}_c^T \underline{P} + \underline{Q} = 0 \quad (49)$$

As it follows from Theorem 1, the sufficient condition for the H_∞ tracking performance existence for the nonlinear system with adaptive fuzzy control law (37)-(39) is that the solution \underline{P} of (40) must be positive definite and symmetric. It can be shown that in order to achieve this requirement the following condition must be satisfied [12]

$$2\rho^2 \geq r \quad (50)$$

i.e., if the inequality (50) is satisfied, then for the nonlinear system (1) the H_∞ tracking performance with the prescribed attenuation level ρ can always be achieved via the adaptive fuzzy control (37)-(39). In general, as ρ is decreased r must be decreased in order to satisfy the inequality (50). However, (37) implies that the control variable u_a must be increased to attenuate w to the desired level ρ . Thus, there is a tradeoff between the H_∞ performance and the control magnitude.

4 Simulation Example

Example 1

The above described adaptive fuzzy control algorithm will now be evaluated using the inverted pendulum system depicted in Fig. 1.

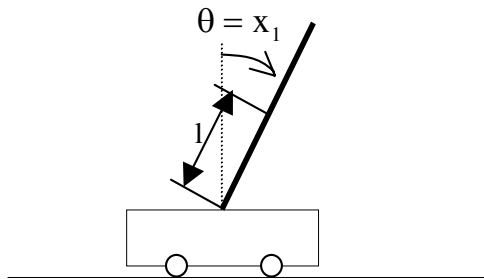


Figure 1
The inverted pendulum system

Let $x_1 = \theta$ and $x_2 = \dot{\theta}$. The dynamic equation of the inverted pendulum is given by [6]

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \frac{g \sin x_1 - \frac{mlx_2^2 \cos(x_1) \sin(x_1)}{m_c + m}}{l \left(\frac{4}{3} - \frac{m \cos^2(x_1)}{m_c + m} \right)} \\ &\quad + \frac{\cos(x_1)}{m_c + m} u_c + d \end{aligned} \quad (51)$$

$$y = x_1$$

where g is the acceleration due to gravity, m_c denotes the mass of the cart, m is the mass of the pole, l is the half-length of the pole, the force u_c represents the control signal and d is the external disturbance. In simulations following parameter values are used: $m_c = 1\text{Kg}$, $m = 0.1\text{Kg}$ and $l = 0.5\text{m}$. The

reference signal is assumed to be $y_r(t) = (\pi/30)\sin(t)$ and an external disturbance $d(t) = 0.1\sin(t)$.

If we require

$$|\underline{x}| \leq \frac{\pi}{6}, |u| \leq 180 \quad (52)$$

and substitute the functions $\sin(\cdot)$ and $\cos(\cdot)$ by their bounds, we can determine the bounds

$$f^M(x_1, x_2) = 15.78 + 0.366x_2^2 \quad (53)$$

$$g^M(x_1, x_2) = 1.46, g_m(x_1, x_2) = 1.12 \quad (54)$$

$k_1 = 2$, $k_2 = 1$ and $\mathbf{Q} = \text{diag}(10,10)$ are set. In order to simplify further calculations $r = 2\rho^2$ is chosen. Then the algebraic Riccati equation solution is

$$\mathbf{P} = \begin{bmatrix} 15 & 5 \\ 5 & 5 \end{bmatrix} \text{ and } \lambda_{\min}(\mathbf{P}) = 2.93. \text{ Five Gaussian membership functions for}$$

both x_1 and x_2 ($i=1,2$) are selected to cover the whole universe of discourse

$$\mu_{F_1^1}(x_i) = \exp\left(-\left(\frac{x_i - \pi/6}{\pi/24}\right)^2\right) \quad (55)$$

$$\mu_{F_1^2}(x_i) = \exp\left(-\left(\frac{x_i - \pi/12}{\pi/24}\right)^2\right) \quad (56)$$

$$\mu_{F_1^3}(x_i) = \exp\left(-\left(\frac{x_i}{\pi/24}\right)^2\right) \quad (57)$$

$$\mu_{F_1^4}(x_i) = \exp\left(-\left(\frac{x_i + \pi/12}{\pi/24}\right)^2\right) \quad (58)$$

$$\mu_{F_1^5}(x_i) = \exp\left(-\left(\frac{x_i + \pi/6}{\pi/24}\right)^2\right) \quad (59)$$

Using the method of trial and errors $\gamma_f = 50$ and $\gamma_g = 1$ are chosen. The pendulum initial position is chosen as far as possible ($\theta(0) = x_1 = \pi/12$) to emphasize the efficiency of our algorithm.

Two cases have been considered in order to show the influence of the linguistic rules incorporation into the control law:

Case one: the initial values of $\underline{\theta}_f$ and $\underline{\theta}_g$ are chosen arbitrarily.

Case two: the initial values of $\underline{\theta}_f$ and $\underline{\theta}_g$ are deduced from the fuzzy rules describing the system dynamic behavior. For example, if we consider the unforced system, i.e. $u_c = 0$, the acceleration is equal to $f(x_1, x_2)$. So intuitively we can state:

“The bigger is x_1 , the larger is $f(x_1, x_2)$ ”.

Transforming this fuzzy information into a fuzzy rule we obtain

$R_f^{(1)}$: IF x_1 is F_1^5 and x_2 is F_2^5 , THEN $f(x_1, x_2)$ is “Positive Big”

where “Positive Big” is a fuzzy set whose membership function is $\mu_{F_1^i}(x_i)$ given by (55)-(59). The acceleration is proportional to the gravity, i.e. $f(x_1, x_2) \cong \alpha \sin(x_1)$, where α is a constant. As $f(x_1, x_2)$ achieves its maximum at $x_1 = \pi/2$, using (53) we obtain $\alpha \cong 16$. The resulting set of 25 fuzzy rules characterizing $f(x_1, x_2)$ is given in Tab. 1.

Table 1
Linguistic rules for $f(x_1, x_2)$

$f(x_1, x_2)$		x_1					
		F_1^1	F_1^2	F_1^3	F_1^4	F_1^5	
		$-\frac{\pi}{6}$	$-\frac{\pi}{12}$	0	$\frac{\pi}{12}$	$\frac{\pi}{6}$	
x_2	F_2^1	$-\frac{\pi}{6}$	-8	-4	0	4	8
	F_2^2	$-\frac{\pi}{12}$	-8	-4	0	4	8
	F_2^3	0	-8	-4	0	4	8
	F_2^4	$\frac{\pi}{12}$	-8	-4	0	4	8
	F_2^5	$\frac{\pi}{6}$	-8	-4	0	4	8

Now the following observation is used to determine the fuzzy rules for $g(x_1, x_2)$:

“The smaller is x_1 , the larger is $g(x_1, x_2)$ ”.

Similarly to the case of $f(x_1, x_2)$ and based on the bounds (53)-(54) this observation can be quantified into the 25 fuzzy rules summarized in Tab. 2.

Table 2
Linguistic rules for $g(x_1, x_2)$

$g(x_1, x_2)$		x_1					
		F_1^1	F_1^2	F_1^3	F_1^4	F_1^5	
		$-\frac{\pi}{6}$	$-\frac{\pi}{12}$	0	$\frac{\pi}{12}$	$\frac{\pi}{6}$	
x_2	F_2^1	$-\frac{\pi}{6}$	1.26	1.36	1.46	1.36	1.26
	F_2^2	$-\frac{\pi}{12}$	1.26	1.36	1.46	1.36	1.26
	F_2^3	0	1.26	1.36	1.46	1.36	1.26
	F_2^4	$\frac{\pi}{12}$	1.26	1.36	1.46	1.36	1.26
	F_2^5	$\frac{\pi}{6}$	1.26	1.36	1.46	1.36	1.26

To obtain the same tracking performances the attenuation level ρ is equal to 0.2 in the first case and to 0.8 in the second one.

The tracking performance of both cases for a sinusoidal trajectory is illustrated in Fig. 2.

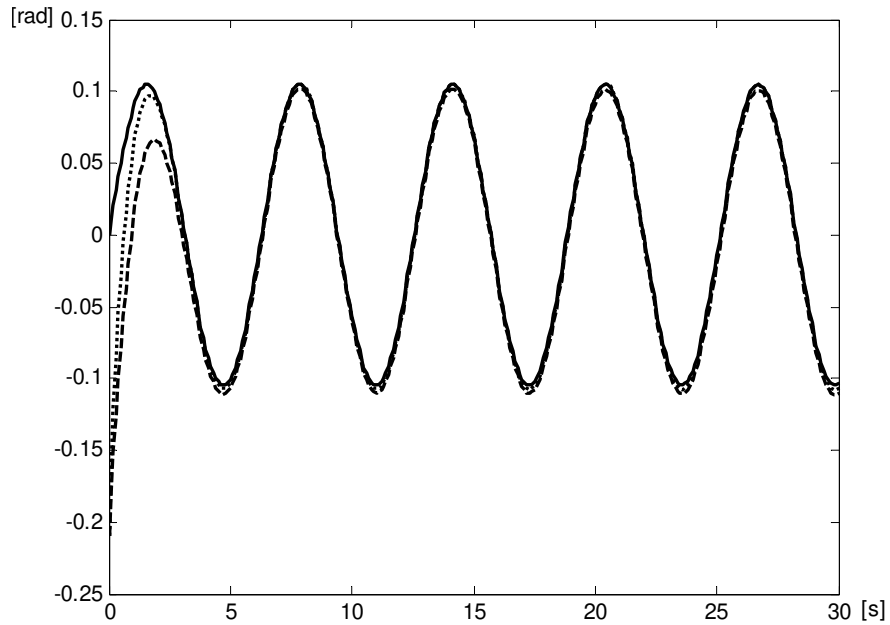


Figure 2

The state X_1 in case 1(dashed line), in case 2 (dotted line) and desired value

$$y_r(t) \text{ (solid line) for } \underline{x}(0) = (\pi/12, 0)^T$$

Example 2

In this example, we apply the adaptive fuzzy controller to the system

$$y'' + \frac{1}{0.25 + y} y' + 1.7y - 0.5u = 0 \quad (60)$$

Define six fuzzy sets over interval $\langle -10, 10 \rangle$ with labels N3, N2, N1, P1, P2, P3. The membership functions are

$$\mu_{N3}(x) = \frac{1}{1 + e^{5(x+2)}} \quad (61)$$

$$\mu_{N2}(x) = \frac{1}{e^{(x+1.5)^2}} \quad (62)$$

$$\mu_{N1}(x) = \frac{1}{e^{(x+0.5)^2}} \quad (63)$$

$$\mu_{P_1}(x) = \frac{1}{e^{(x-0.5)^2}} \quad (64)$$

$$\mu_{P_2}(x) = \frac{1}{e^{(x-1.5)^2}} \quad (65)$$

$$\mu_{P_3}(x) = \frac{1}{1 + e^{-5(x-2)}} \quad (66)$$

The reference model is assumed to be

$$M(s) = \frac{1}{s^2 + 2s + 1} \quad (67)$$

and the reference signal is the series of jumps with variant magnitude.

We choose $\mathbf{P} = \begin{bmatrix} 50 & 30 \\ 30 & 20 \end{bmatrix}$, $k_1 = 2$, $k_2 = 1$, and $\lambda_{\min}(\mathbf{P}) = 1.52$. To satisfy

the constraint related to $|\underline{x}|$ we choose $\rho = 0.01$.

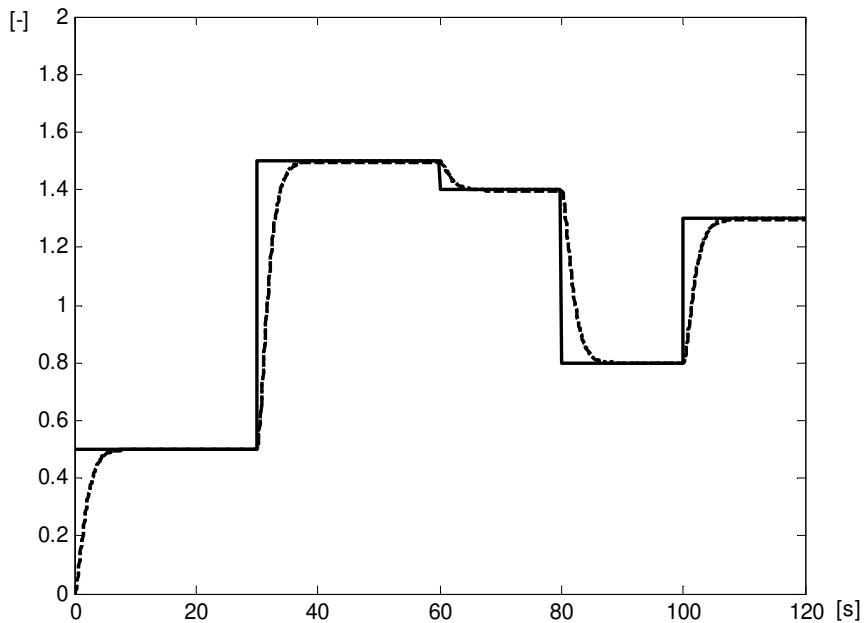


Figure 3

The state x_1 (dashed line), its desired reference model value $y_m(t)$ (dotted solid line) and reference signal (solid line)

At 75th second of simulation the system (60) was switched to another system

$$y''' + 5y'' + \left[\frac{1}{(0.25 + y)^2} - 1.7 \right] y' + y - 5u = 0 \quad (68)$$

All initial states have been set to zero $y(0) = y'(0) = y''(0) = y'''(0) = 0$.

As it can be seen from Fig. 3, the simulation results confirm good adaptation capability of the proposed control system. The system dynamic changes are in particular manifested by changes of control input signal (Fig. 4).

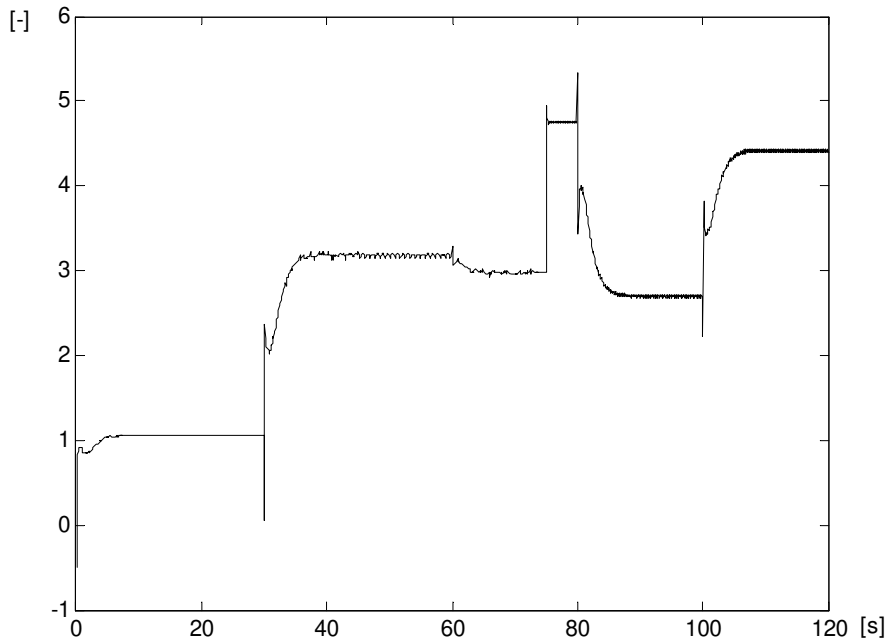


Figure 4
Control signal

Conclusions

In this paper the adaptive fuzzy controller has been proposed for the class of nonlinear systems subject to large uncertainties or to unknown variations in the parameters and the structure of the plant.

The proposed adaptive control scheme has involved both fuzzy systems and H_∞ control. The adaptive fuzzy systems can be considered as a rough tuning control for approximation of the nonlinear system and the H_∞ control can be considered as a fine-tuning control used to filter the approximation errors and external

disturbances. Therefore, the proposed adaptive algorithm will be useful for the unknown (or uncertain) nonlinear system control design. The simulation results show that approximation errors and external disturbances can be successfully attenuated using the proposed control design method within a desired attenuation level, i.e. H_∞ tracking performance is achieved.

Further work is under investigation to apply the proposed robust adaptive algorithm to multi input multi output (MIMO) systems.

References

- [1] K. M. Passino, S. Yurkovich, "Fuzzy Control." Addison-Wesley, California 1998
- [2] A. Isidori, "Nonlinear Control Systems," Berlin, Springer-Verlag, 1989
- [3] J. J. E. Slotine, W. Li, "Composite Adaptive Control of Robot Manipulators," *Automatica*, Vol. 25, pp. 509-519, 1991
- [4] D. G. Taylor, P. V. Kokotovic, R. Marino, I. Kanellakopoulos, "Adaptive Regulation of Nonlinear Systems with Unmodeled Dynamics," *IEEE Trans. on Auto. Contr.*, Vol. 34, pp. 405-412, 1989
- [5] S. S. Sastry, A. Isidory, "Adaptive Control of Linearizable Systems," *IEEE Trans. on Auto. Contr.*, Vol. 34, pp. 1123-1131, 1989
- [6] L. X. Wang, "Stable Adaptive Fuzzy Controllers with Application to Inverted Pendulum Tracking," *IEEE Trans. on Syst., Man and Cybernetics-part B*, Vol. 26, pp. 677-691, 1996
- [7] L. X. Wang, J .M. Mendel, "Fuzzy Basis Functions, Universal Approximation, and Orthogonal Least-Squares Learning", *IEEE Trans. on Neural Networks*, Vol. 3, No. 5, September 1992
- [8] J. T. Spooner, K. M. Passino, "Stable Adaptive Control Using Fuzzy Systems and Neural Networks," *IEEE Trans. Fuzzy Syst.*, Vol. 4, August 1996
- [9] H. Han, Chun-Yi Su, Y. Stepanenko, "Adaptive Control of a Class of Nonlinear Systems with Nonlinearly Parametrized Fuzzy Approximators," *IEEE Trans. Fuzzy Syst.*, Vol. 9, April 2001
- [10] Ch. H. Wang, H. L. Liu, T. Ch. Lin, "Direct Adaptive Fuzzy-Neural Control with State Observer and Supervisory Controller for Unknown Nonlinear Dynamical Systems", *IEEE Trans. on Fuzzy Systems*, Vol. 10, No. 1, February 2002
- [11] W.-Y. Wang, Y.-G. Leu, C. C. Hsu, "Robust Adaptive Fuzzy-Neural Control of Nonlinear Dynamical Systems Using Generalized Projection Update Law and Variable Structure Controller," *IEEE Trans. on Syst., Man and Cybernetics-part B*, Vol. 31, pp. 140-147, 2001

-
- [12] M. Kratmüller, „The Adaptive Control of Nonlinear Systems Using the T-S-K Fuzzy Logic,” *Acta Polytechnica Hungarica, Journal of Applied Sciences at Budapest Tech, Hungary*, Volume 6, Issue Number 2, 2009, pp. 5-16, ISSN 1785-8860
- [13] T. Basar, P. Bernhard, “ H_∞ Optimal Control and Related Minimax Problems: A Dynamic Game Approach,” Birkhäuser, Berlin, Germany 1991
- [14] J. W. Helton, O. Merino, “Classical Control Using H_∞ Methods,” SIAM Philadelphia, 1998
- [15] L. X. Wang, “Adaptive Fuzzy Systems and Control, Design and Stability Analysis,” PTR Prentice Hall, 1994
- [16] T. J. Koo, Analysis of a Class of Fuzzy Controllers, in Proc. 1st Asian Fuzzy Systems Sump., Singapore, Nov. 1993
- [17] M. Kratmüller, J. Murgaš, “Priame adaptívne riadenie s fuzzy prístupom,” *Kybernetika a informatika, Trebišov 2002* (in Slovak)
- [18] B. S. Chen, T. S. Lee, J. H. Feng, “A Nonlinear H_∞ Control Design in Robotic Systems under Parameter Perturbation and External Disturbance,” *INT. J. CONTROL*, Vol. 59, No. 2, 439-461, 1994

How Safe the Human-Robot Coexistence Is? Theoretical Presentation

Olesya Ogorodnikova

Department of Mechatronics, Optics and Informational Engineering
Budapest University of Technology and Economics
Műegyetem rkp. 3-9, H-1111 Budapest, Hungary
E-mail: olessia@git.bme.hu

Abstract: It is evident that industrial robots are able to generate forces high enough to injure a human. To prevent this, robots have to work within a restricted space that includes the entire region reachable by any part of the robot. However, more and more robot applications require human intervention due to superior abilities for some tasks performance. In this paper we introduce danger/safety indices which indicate a level of the risk during interaction with robots, which are based on a robot's critical characteristics and on a human's physical and mental constrains. Collision model for a 1 DOF robot and "human" was developed. Case study with further simulations was provided for the PUMA 560 robot.

Keywords: robotics, safety, human-robot interaction, danger index

1 Introduction

Most safety standards require an installation of the safeguarding systems, so any access to the hazard (robot work space) is prevented, or the cause of hazard is removed without requiring specific conscious action by the person. The prescribed action to be taken by the robot system upon detecting an intrusion into the safeguarding space is to remove all drive power and all other energy sources. Thus, robots must be surrounded by the safeguarding space and production must be designed to allow the maximum number of tasks to be performed with personnel outside the safeguarding space. However, this approach is not applicable for the new tendency in robotics where humans and robots interact in unstructured space and where their working zones are overlapped (social robotics, collaborative tasks, etc.) [1-3]. A typical situation in industrial robotics where a human operator can be hit, trapped between the safety equipments and the robot parts is during maintenance, teaching or collaboration [4], [5]. To avoid or minimize the severity of injury we should keep the risk level at a minimum.

Concerning injuries caused by robots, only very little data or literature is available. In [6] the United Auto Workers (UAW) union published a report which provides raw data on various injuries related to robot operations. There are many types of injuries which could potentially occur during the interaction between a human and a manipulator. These include cuts or abrasions, which might result from contact with a sharp or abrasive surface, as well as more serious injuries including bone fracture which could result from manipulator pinch points or direct crush loads. However, when a human operator works near a robot, the most dangerous accident is the potential impact with large loads that may cause serious injury or even death.

Therefore, the danger criterion should be constructed from measures that contribute to reducing the impact force in case of unexpected human-robot impact, as well as reducing the likelihood of the impact itself. Concerning to this issue we introduced a new criteria (danger index) which is based on the critical measures of impact forces, accelerations and distances to reduce a probability of the dangerous collision.

2 Related Work

A number of standard indices of injury severity have been developed. Some of them attempt to relate resulting head acceleration to the severity and likelihood of injury [7], [8], [9], [10]. The basis of these measures is the Wayne State University Tolerance Curve (WSTC) (See Fig. 1) which relates acceleration and duration to the likelihood of severe brain injury.

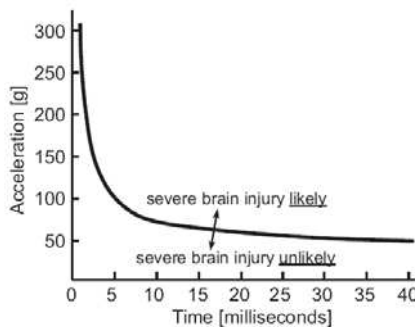


Figure 1

The Wayne State tolerance curve [12]

To evaluate the potential for serious injury due to impact an empirical formula was developed by the automotive industry to correlate head acceleration to injury severity known as the Head Injury Criteria (HIC) [11], which is computed as the

maximum integral of the resultant acceleration of the centre of mass of the head during the crash (1).

$$HIC = \Delta t \left(\frac{1}{\Delta t} \int_{t_1}^{t_2} a_h dt \right)^{2.5}, \Delta t = t_2 - t_1 \quad (1)$$

Where a_h is the resulting acceleration of the human head and Δt is a period of impact that should not be more than 15 ms.

Prasad and Mertz [12] introduced a set of curves which statistically relates measured HIC values to the severity and likelihood of a head injury. Using these curves, in combination with evaluated HIC values, it is possible to define the level of an injury resulting from a given head acceleration time history. The resulting injury indices can be also used to judge the severity of the injury with further consultation of biomechanical expertise, like e.g. the so called Abbreviated Injury Scale (AIS) [13]. Figure 2 illustrates an exponential correlation between AIS and HIC criterion, which evaluation is based on post mortal experiments. It is seen that from a certain value of AIS (1.6) the HIC rises drastically. The HIC of 250 is correlated to the AIS1+ value where injuries to the human are negligible.

The HIC is a commonly used frontal impact criteria that has been used for decades to assess the level of head injury risk in frontal collisions. A HIC of 1000 is conventionally considered to represent the threshold where linear skull fractures normally begin to appear. According to this assumption a head can sustain acceleration more than 90 g. However, for the lateral or transversal impacts this value can result in severe injuries especially if this acceleration was caused by collision with a rigid surface (manipulator arm). Therefore, more experiments have to be provided and more restricted boundaries have to be introduced.

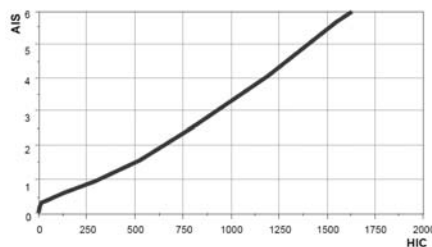


Figure 2

The AIS – HIC dependency curve [13]

Most research related to the HIC criteria were based on the automobile crash-testing results defined this criterion as an impact involving a collision of the head with another solid object at appreciable velocity. This situation is generally characterized by large linear accelerations and small angular accelerations during the impact phase.

In the work [14] Head Injury Criterion was evaluated for robot masses up to 500 kg with the linear velocities (0.2, 0.5, 1.0, 1.5, 2.0, 2.5, 3) m/s. Results indicated that at some point increasing robot mass does not result in a higher HIC. Moreover, according to this criterion, no robot, whatever mass it had, became dangerous at the operating speed up to 2 m/s as long as the time of impact was less than 36 ms. In this research typical severity indices, established in the automobile industry, were claimed to be not applicable for a human-robot interaction domain in view of the much lower operating velocities.

In another study [15] Manipulator safety index (MSI) was introduced, which evaluation was also based on the HIC criterion, where a head acceleration was computed from a human robot a collision model. The resulted index depends on the manipulator's effective inertia, interface stiffness, and initial robot velocity mappings. However, in graphical representations of this analysis the velocity and the stiffness characteristics become dominant while the inertial weight - negligible. Moreover, in view of the conflict of some parameters (low stiffness with high inertia configuration) this index can not reflect the real hazard caused by the robot under its certain configurations. In spite of the fact that the advantage of the HIC criteria application in the robotic field is questionable at some extent in this research evaluations were also based on this criteria.

3 An Introduction to a Danger Severity Evaluation

Since it is not feasible to adequately treat all different contact types of injuries in this work only blunt contacts were considered. To develop a quantitative measure which relates the severity and likelihood of injury to the physical characteristics of a given manipulator factors as force, acceleration and distance were taken into account. In the case of mechanical injury at a collision accident, the severity of an injury mostly depends on impact force and the likelihood depends on the distance to impact area before collision. In turn, an impact force mostly depends on robots physical characteristics, specific configurations, approaching speed, direction, and the contact duration. [16], [17] Among the minor factors that contribute to the Index are diverse robot tasks, failure rates, presence of any safety features, shape of the instrument, joint compliance, control methods, etc. Moreover, the severity of impact will also depend on the human factor [18]. For instance, characteristics as age, sex, weight will change personnel physical and mental hazard perception as well as a reaction on it. In this paper human physical constrains were considered to establish the boundaries on the robot performance, assuming that a physical contact may occur. Critical characteristics were obtained from the biomechanical injury/pain tolerance estimations, acquired experimentally in the works [7], [19]. These results were approximated and the mean values were used.

The proposed generalized form of the danger index consists of a linear combination of qualities that take into account the relevant distance ($Did(t)$), the contact force ($Di_f(t)$) and the human head acceleration factors ($Di_a(t)$) (2). The sum of these indices with their corresponding weights has to be less or equal to one for a safe human robot interaction:

$$DI = \alpha_d Di_d(t) + \alpha_f Di_f(t) + \alpha_a Di_a(t) \leq 1$$

$$Di_d(t) = \frac{L_c}{L_i}, \quad Di_f(t) = \frac{F_i}{F_c}, \quad Di_a(t_c) = \frac{a_i}{a_c} \quad (2)$$

Where F_i is an actual value of the force exerted by a manipulator, i.e. a producible impact force of the robot, F_c is a critical, admissible force, that doesn't cause serious injury to a human at a collision body part. L_i is an actual distance measured from the visual or sensory monitoring system, L_c is a distance that robot overpass after stopping signal was sent to the robot control. This stopping distance mainly depends on the actual robot speed v_i and its load. Parameter a_i is an acceleration of a head measured after collision with manipulator that is compared with a critical one a_c obtained from the AIS scaling.

All indices are time dependent. An acceleration related index is examined under the condition when the head acceleration (or other body part) achieves the maximum value. This occurs at the minimum or critical time interval Δt . [20] Coefficients α_d , α_f and α_a are weights of the distance, force and acceleration terms respectively. The indices evaluation and the corresponding weights distribution is based on the initial task description, risk assessment results and information available during analysis. For instance, for collaborative tasks in close vicinity distance factor is not important since the distance is negligible or even contact between human and robot is possible. However, if a robot effective mass and, as a consequence, exerted force at some configurations is greater than the admissible value human can be injured. In this case the force related danger index plays the dominant role. On the other hand, if a robot is performing task in the automatic regime with the maximum (optimal for the task) characteristics, it is essential to keep the safe distance to avoid the likelihood of impact under these conditions, i.e. the weight for the distance related index will be under consideration.

3.1 Distance Related Danger Index

Sufficient distance provides with time to reduce impact force by braking actions to avert the collision. Thereby, keeping that distance can be a criterion for a danger evaluation. To compute this value we should know mutual robot, personnel approaching speed and the time needed to stop all movements.

A minimum distance to hazard L_c (3) depends on a robot's operational speed, a sensory system reaction time, control system response time and robot's braking characteristics. Time T_i expresses a robot's stopping time that varies depending on

the applied drivers stopping category, braking system idle time and the safety system response time (if the safety distance is controlled by the external present sensing device). [21] It was assumed that in (3), (4) acceleration (deceleration) is a constant value.

$$L_c = v_i \times T_i \quad (3)$$

$$L_i = (v_i + v_h)t - at^2 / 2 > L_c \quad (4)$$

$$Di_d = (v_i \times T_i) / ((v_i + v_h)t - at^2 / 2) < 1 \quad (5)$$

When the distance to contact is sufficient (4) we have time to decelerate the robot and avoid the undesired impact. When the speed of a robot can be reduced with some deceleration to the condition when the contact with a human becomes not likely, the distance is claimed to be safe. At this distance robot can move at its normal operational speed and in the case of a safety distance violations, it decelerates or cease all movements. In (4), (5) v_h is a human average walking, hazard approaching speed. According to the human factor analysis its mean value is 1.6 m/s. v_i is a robot operational speed, t is a time scale. At the distance L_i robot is fulfilling its task at the max speed or at the speed needed for the effective task performance till there is no human entering the monitoring area. As soon as a no authorized access to this zone has been recognized robot's speed is decreases with an acceleration a . If the critical distance L_c is overrun (or near to be), a robot is forced to stop. This situation occurs if human continues to move toward the robot in spite of the warnings, or if the robot does not have enough time to decelerate to the speed established as a "safe" for the current distance at the time t .

Therefore, the distance related danger index is evaluated basing on the relation between the critical and the current distances, where the later should be kept always greater than the critical one to avoid undesirable contact. This danger index formulation is represented in (5). Danger index can be displayed as a circle (sphere) which radii corresponds to the value 1. All characteristics inside this circle will comply with the safety requirements, exceeded values will require appropriate danger reduction procedures.

For our analysis robot speeds v_i were set in the interval [0, 0.2, 0.7, 1, 1.5, 2, 2.9] m/s. Time T_i with respect to the speed v_i was chosen according to the experimental results provided in the work [22] From the relation in (5) we define the time interval where in compliance with the danger index analysis this function should be less than zero (6). Fulfillment of this condition decreases the probability of the human robot contact since the minimum distance between them is provided by the danger index control.

$$f(t) = at^2 / 2 - (v_{ir} + v_h)t + v_{ir}T_i < 0 \quad (6)$$

From Fig. 3 we can see that the requirements are met within the time interval [t_1 , t_2]. Graphical representations in Fig. 4 (a) indicate the minimum required

deceleration values for the speed range: 0.25, 0.6 and 1 m/s. It is evident that lower velocities need less time to decelerate. Integrating human walking speed in the danger index formulation system has to apply greater accelerations to convey with the safety (danger index) requirements. (See Fig. 4 (b))

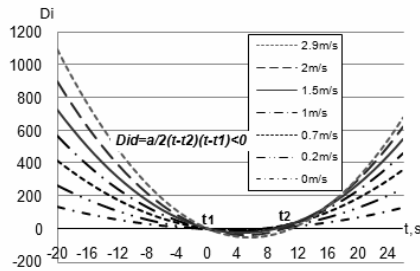
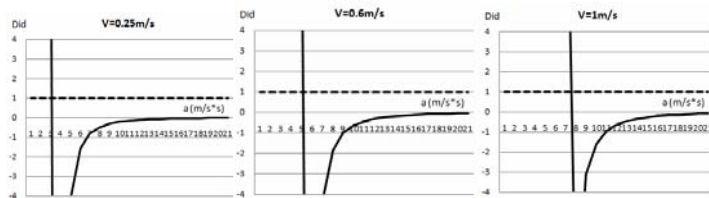
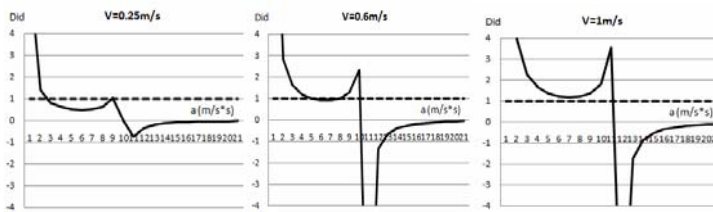


Figure 3
The distance related danger index function



(a)



(b)

Figure 4

An acceleration dependent danger index representation: with human movement (a) and without (b) in consideration ($T_i=0,5$ s, linear velocities v : 0.25, 0.6, 1 m/s)

3.2 Force Related Danger Index

The force F as well as acceleration a in general can be defined as a function of different influencing elements, i.e.: approaching velocity, robot effective mass, inertia, stiffness, kinetic energy, etc. In this study an effect of the manipulator arm effective mass and approaching speed is investigated. According to the Newton's theory the impact force depends on the robot (here) acceleration or speed at the moment of collision, therefore, both characteristics will be considered in the analysis. In general, the second Newton's law formalization provides with description of the linear motion, where applied force F depends on the mass m moving with acceleration a . This formulation can be also described in terms of the linear momentum mv where a rate of change of the linear momentum is equal to the applied forces. With a reference to a collision model (will be discussed later) we yield (8):

$$\frac{m_u v_0 - m_u v'}{\Delta t} = F_u \quad (8)$$

Where Δt is a time of the collision duration, v' is a velocity after impact, m_u is a scalar value of the mass at the direction u , F_u is a resulted force at the same direction. At some conditions the value of this force can become infinite or very large. This situation is very dangerous especially when the human is under the risk of impact. Therefore, to establish tolerable boundaries on the exerted force magnitudes is an issue that has to be investigated. One of the possible solutions is to introduce a danger index (9) that would indicate an admissible level of the controlled parameter.

$$Di_f = \frac{F_i}{F_c} \leq 1 \quad \left(Di_a = \frac{a_i}{a_c} \leq 1 \right) \quad (9)$$

In the force related danger evaluation the producible force F_i should be compared with the critical one F_c , which is maximum "safe" value, and established basing on the largest force magnitude that does not cause serious injury or pain to a human (here). In the course of the injury limit evaluation the most vulnerable part of a human body was considered, head.

Similarly, with respect to Newton's law, acceleration related index can be yielded. However, it was also decided to investigate HIC criteria and its applicability for a HRI field, thus, two critical acceleration will be used for a acceleration related danger estimation, based on Newton's law and on HIC index.

According to the studies provided in [22] the level of injury can be measured on the basis of the head human skull bone fractures, however, the threshold of the fracture highly depends on the contact area. For instance, the fracture force of the occipital bone is estimated 6.41 KN, while the fracture force of the maxilla bone was measured of only 0.66 KN. On the other hand, considering that analysis is provided for friendly human-robot interactions even any causes of pain should be

avoided. Since this characteristic also depends on the area of impact, each body part can be considered separately.

Therefore, it was determined for the most critical estimations (pain tolerance) apply more restrictive tolerance limits on the robot exerted forces. There have been few reports discussing the human pain tolerance limits when static or dynamic stimuli are applied to the whole body. For this study a critical force value causing pain was derived from the analysis provided in the work [19], where somatic pain tolerance is investigated. Parameters of the pain tolerance were acquired from a human response on the applied mechanical stimuli. For instance, for the parts under the most frequent exposure to the hazard (hand, arm, back and head) the critical force was found as 140 N, 180 N, 240 N and 130 N respectively. In this analysis, for the further evaluations an effect of a head impact will be investigated. Thus, the most restrictive danger criteria will be based on the force equal 130 N.

4 Manipulator Effective Mass Formulation

For a multi-link manipulator the effective mass at the direction of impact is changing with each robot configuration. We consider the impact itself in the operational space of a manipulator, therefore, the mass and inertial properties have to be evaluated in that space. Since the mass properties of a manipulator are generally expressed with respect to its motion in joint space the transformation method should be introduced.

The manipulator's dynamic model in the joint (10) and operational spaces (11) is described in [23].

$$M(q)\ddot{q} + v(q, \dot{q}) + g(q) = \tau \quad (10)$$

$$M_x \ddot{x} + v_x(x, \dot{x}) + g_x(x) = F \quad (11)$$

Here $M(q)$ is $n \times n$ joint and M_x is end effector kinetic energy matrices, $v(q, \dot{q})$, $v_x(x, \dot{x})$ are the vectors of centrifugal and corioles forces, $g(q)$ is the vector of gravity, τ , F are the generalized vectors of joint and end effector force respectively. The relation between two matrices can be expressed as in (12).

$$M_x(q) = (J(q)M^{-1}(q)J^T(q))^{-1} \quad (12)$$

Where $J(q)$ is the basic Jacobian associated with the end-effector linear and angular velocities and $M(q)$ is a symmetric positive defined mass matrix. Assuming that impact occurs within a robot's transition movement (close distance collision), $J(q)$ is equal to $J_v(q)$ (Jacobian matrix associated with the linear velocity of the end effector). If an impact occurs when the end-effector is moving along an

arbitrary direction, a kinetic energy matrix in this case is a scalar (m) representing the mass perceived at the end effector (point of impact) in response to the application of a force (F) along this direction (13) (See Fig. 5).

$$\frac{1}{m_u} = J_{v_u}^T(q) M_v^{-1}(q) J_{v_u}(q); \quad J_{v_u}(q) = u^T J_v(q) \quad (13)$$

To evaluate the effective mass at the direction of impact the mass matrix $M_v(q)$ should be diagonalized in order to avoid the effect of coupling between its elements. One of the methods that can be introduced is the eigenvectors (V) and eigenvalues (λ) determination with an ellipsoidal geometrical representation of the mass matrix properties as it is shown in (14). This representation provides a description of the square roots of the effective mass properties (eigenvalues) in the arbitrary directions (eigenvectors) [24].

$$\left(\frac{x}{\sqrt{\lambda_1}}\right)^2 + \left(\frac{y}{\sqrt{\lambda_2}}\right)^2 + \left(\frac{z}{\sqrt{\lambda_3}}\right)^2 = 1 \quad (14)$$

The eigenvalues and eigenvectors associated with the matrix $M_v(q)$ or its inverse provide a useful characterization of the bounds on the magnitude of the mass properties. The eigenvectors of this matrix define the principal directions of the ellipsoid and the inverse of the square roots of the eigenvalues indicate the corresponding equatorial radii. Moreover, by identification of the maximum eigenvalues (eigenvectors) characteristics (15), it is possible to assess the extent of the manipulator actual configuration danger and establish corresponding boundaries in compliance with safety requirements and danger criteria of the task (16). (See Fig. 5)

$$\frac{1}{m_{\max}(\lambda_{\max})} = V^T(\lambda_{\max}) M_v^{-1}(q) V(\lambda_{\max}) \quad (15)$$

$$\frac{1}{m_c(Di)} = u_c^T(Di) M_v^{-1}(q) u_c(Di) \quad (16)$$

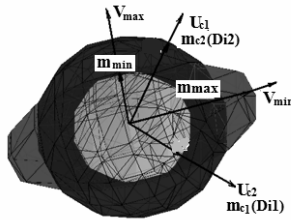


Figure 5

Effective mass ellipsoid with two intersecting danger index (Di1, Di2) spheres

5 Collision Modeling

For more precise danger indices analysis we refer to the dynamic simulation of the impact which is based on the one DOF mass-spring collision model (See Fig. 6). An assumed dynamic model is described in the equations of motion in (17):

$$M_r a_r + Ke(x_r(t) - x_h(t)) = 0 \quad (17)$$

In here M_r and a_r are manipulator arm effective mass acting in the direction of impact and its deceleration value after collision respectively, Ke is a measured effective stiffness, difference in displacements x_r and x_h describes a robot and a human (head) mutual allocation after impact.

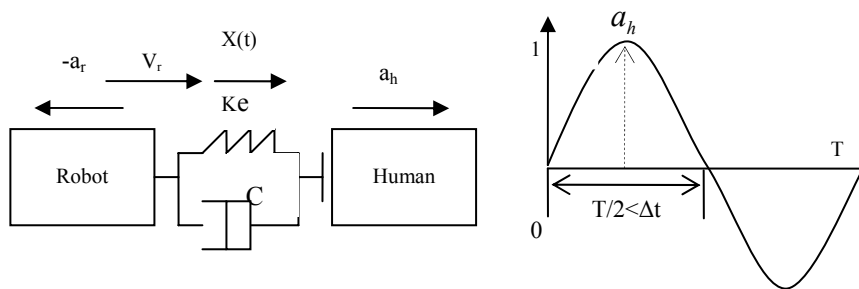


Figure 6
Mass-spring collision model [15]

In this assumption a mass M of the manipulator is an effective mass that reflects the inertial manipulator properties at the point of impact. The real value of the acceleration a_h and the period of impact can be found from the relations below, assuming that the impact occurs with a maximum spring compression $x(t)_{\max}$ defined from (18):

$$\frac{1}{2} M v_{or}^2 = \frac{1}{2} K x^2(t)_{\max}$$

$$x(t)_{\max} = \sqrt{\frac{M v_{or}^2}{K_e}} \quad (18)$$

$$x(t) = x(t)_{\max} \cos(\omega_n t) \quad (19)$$

Setting first derivative of the time dependant generalized form equation of motion (19) equal to 0 the impact period when a head is exposed to a maximum acceleration a_h can be evaluated:

$$\dot{x}(t) = x(t)_{\max} \sin(\omega_n t) = 0 \Rightarrow t(a_{\max}) = \frac{\pi}{\omega_n} = \frac{T}{2} \quad (20)$$

Here ω_n is a natural frequency of the oscillation after impact. For stiff surfaces, as a robot is, this period is assumed to be less than duration of the impact ($T/2 \leq \Delta t$; $\Delta t = 0,015ms$). Further, considering provided above measures, a manipulator and a head accelerations after impact had been estimated:

$$a_r = v_{or} \omega_n \sqrt{\frac{M}{M+m}} \cos(\omega_n t) \quad (21)$$

$$a_h = \frac{M}{m} v_{or} \omega_n \sqrt{\frac{M}{M+m}} \cos(\omega_n t) \quad (22)$$

Where $\cos(\omega_n t) = 1$ if $-T/2 < t < T/2$

Accelerations can be also computed for the mass-spring-damper system, which behavior depends on the natural damping ratio ζ_n (23). The system is critically damped when $\zeta_n = 1$, over damped if $\zeta_n > 1$, and oscillatory damped when $\zeta_n < 1$. The equation of motion for this system is shown in (24):

$$\zeta = \frac{C}{2\sqrt{K_e M}} \quad (23)$$

$$x(t) = x(t)_{\max} e^{-\zeta \omega_d t} \cos(\omega_d t)$$

$$\omega_d = \omega_n \sqrt{1 - \zeta^2}, \quad (24)$$

$$T/2 = \frac{\pi}{\omega_n \sqrt{1 - \zeta^2}}$$

Here ω_d is a damped natural frequency, ζ is a damping ratio and C is a friction coefficient.

Consequently, the head acceleration can be found similarly to (22) and expressed as in (25):

$$a_h = \frac{M}{m} v_0 \sqrt{\frac{M}{M+m}} \omega_n e^{-\zeta \omega_d t} (\zeta^2 - 1) \cos(\omega_d t) \quad (25)$$

Where $\cos(\omega_d t) = 1$ if $-T/2 < t < T/2$

However, in a view of the fact that the robot (here) has a very high stiffness, damping ratio will be very small (10^{-4}) and doesn't contribute significantly to a danger index value. Therefore, for the further computations mass spring damping system will not be considered.

Finally, according to estimations provided in (2), (9), knowing acceleration and force critical values we can establish the acceleration and the force related danger indices.

$$Di_{a_h} = \frac{a_h}{a_c} = \frac{\frac{M}{m} v_0 \omega_n \sqrt{\frac{M}{M+m}}}{a_c} \leq 1 \quad (26)$$

$$Di_f = \frac{f_a}{f_c} = \frac{M v_0 \omega_n \sqrt{\frac{M}{M+m}}}{f_c} \leq 1 \quad (27)$$

6 HIC-based Danger Estimation

According to AIS scale a head can sustain quite high accelerations if the loading is relatively short and if the time duration is relatively long.

Table 1 [8] demonstrates a relation between the peak linear head acceleration and the severity injury level. In this research we consider situations where “no” or only “minor” injuries are acceptable. Therefore, according to the AIS scale the threshold for a maximum head acceleration has been established up to 50 g with the assumed impact duration $\Delta t = 15$ ms. These assumptions imply the HIC 265 computed in (26) that is correlated with the AIS1 level.

$$HIC = \Delta t \left[\frac{\Delta V}{g \Delta t} \right]^{2.5} = 0.015 \times [50]^{2.5} \quad (28)$$

Table 1
AIS Head Injury Scale [13]

Peak linear acceleration, g	AIS head injury severity	Injury interpretation
<50	0	No
50-100	1	Minor
100-150	2	Moderate
150-200	3	Serious
200-250	4	Severe
250-300	5	Critical
>300	6	Unsurvivable

By substituting identified accelerations in Ch5 into a HIC criterion formulation (1) we can establish a relation between the AIS scale and the manipulator based collision model as in (29):

$$HIC = \Delta t \left[\frac{1}{g \Delta t} \int_{-T/2}^{T/2} \frac{M}{m} v_0 \omega_n \sqrt{\frac{M}{M+m}} \cos(\omega_n t) dt \right]^{2.5} \quad (29)$$

Taking a define integral from (29), a new HIC criteria that depends on the manipulator's operating characteristics can be yielded as following:

$$HIC = \Delta t \left[\frac{2v_0 \frac{M}{m} \sqrt{\frac{M}{M+m}}}{g\Delta t} \sin(\omega_n \frac{\Delta t}{2}) \right]^{2.5} \quad (30)$$

Where $\sin(\omega_n (\Delta t/2))=1$ for $\Delta t > T/2$

Hence, human head acceleration from HIC index is:

$$a_h(HIC) = \frac{2v_0 \frac{M}{m} \sqrt{\frac{M}{M+m}}}{\Delta t} \quad (31)$$

Furthermore, substituting this expression into (9) a HIC-based acceleration and force related danger criterion can be obtained:

$$Di_{a_h}(HIC) = \frac{a_h}{a_c} = \left(\frac{2v_0 \frac{M}{m} \sqrt{\frac{M}{M+m}}}{\Delta t} \right) / a_c \leq 1 \quad (32)$$

$$Di_f(HIC) = \frac{f_a}{f_c} = \left(\frac{2v_0 M \sqrt{\frac{M}{M+m}}}{\Delta t} \right) / f_c \leq 1 \quad (33)$$

By definition, these two danger evaluations should be equivalent and both can be used independently for the safety level identification.

7 Case Study

To cite an example a PUMA 560 robot was applied (See Fig. 7). First, an effective mass matrix at a given robot configurations: $q_1(0)$, $q_2(0)$, $q_3(0)$, $q_4(0)$, $q_5(90)$, $q_6(10)$ grad was computed. Analysis was provided with two assumptions: the last 3 joints of the robot do not contribute significantly into a kinetic energy matrix of the PUMA robot, therefore, the mass matrix $M(q)$ in joint space has a dimension 3×3 ; the distance before collision is relatively small, thus, the motion of the end effector in the direction of impact is considered translational.

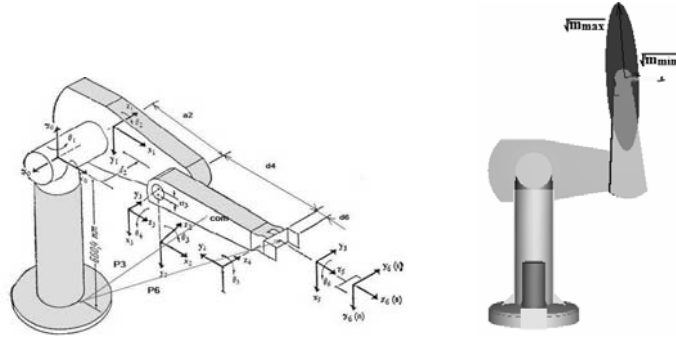


Figure 7
PUMA 560

In Fig. 7 vectors P3 and P6 identify the center of mass of the link 3 and the point of impact respectively:

$$P3 = \begin{bmatrix} p_{3x} \\ p_{3y} \\ p_{3z} \end{bmatrix} = \begin{bmatrix} C_1(a_2C_2 + a_3C_{23}) - d_2S_1 \\ S_1(a_2C_2 + a_3C_{23}) + d_2C_1 \\ -a_3S_{23} - a_2S_2 \end{bmatrix} \quad (34)$$

$$P6 = \begin{bmatrix} p_{6x} \\ p_{6y} \\ p_{6z} \end{bmatrix} = \begin{bmatrix} C_1(d_6(C_{23}C_4S_5 + S_{23}C_5) + S_{23}d_4 + a_3C_{23} + a_2C_2) - S_1(d_6S_4S_5 + d_2) \\ S_1(d_6(C_{23}C_4S_5 + S_{23}C_5) + S_{23}d_4 + a_3C_{23} + a_2C_2) + C_1(d_6S_4S_5 + d_2) \\ d_6(C_{23}C_5 + S_{23}C_4S_5) + C_{23}d_4 - a_3S_{23} - a_2S_2 \end{bmatrix}$$

According to (10) and basing on the evaluations provided in [25] the mass matrix in a joint space was computed according to the assumption in (35) and for the considered joint angles its final numerical form is presented in (36):

$$M(q) = m_1 J_{v1}^T J_{v1} + m_2 J_{v2}^T J_{v2} + m_3 J_{v3}^T J_{v3} + J_{\omega 1}^T I_{C1} J_{\omega 1} + J_{\omega 2}^T I_{C2} J_{\omega 2} + J_{\omega 3}^T I_{C3} J_{\omega 3} \quad (35)$$

$$M(q)^{3 \times 3} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ X & m_{22} & m_{23} \\ X & X & m_{33} \end{bmatrix} = \begin{bmatrix} 4 & -0.1 & -0.1 \\ -0.1 & 6.8 & 0.3 \\ -0.1 & 0.3 & 1.2 \end{bmatrix} \quad (36)$$

To identify the mass matrix in the manipulator operation works space according to formulation in (12), Jacobian of the P6 vector (impact point) is yielded as in (37):

$$Jv_x = \begin{bmatrix} -[S_1(d_6(C_{23}C_4S_5 + S_{23}C_5) + S_{23}d_4 + a_3C_{23} + a_2C_2) + C_1(d_6S_4S_5 + d_2)] \\ C_1(d_6(-S_{23}C_4S_5 + C_{23}C_5) + C_{23}d_4 - a_3C_{23} - a_2S_{23} - a_2S_2) \\ C_1(d_6(-S_{23}C_4S_5 + C_{23}C_5) + C_{23}d_4 - a_3C_{23} - a_2S_{23}) \\ -d_6C_1C_{23}S_4C_5 - S_1d_6C_4S_5 \\ C_1(d_6C_{23}C_4C_5 - S_{23}S_5) - S_1d_6S_4C_5 \end{bmatrix} \quad (37)$$

$$Jv_y = \begin{bmatrix} C_1(d_6(C_{23}C_4S_5 + S_{23}C_5) + S_{23}d_4 + a_5C_{23} + a_2C_2) - S_1(d_6S_4S_5 + d_2) \\ S_1(d_6(-S_{23}C_4S_5 + C_{23}C_5) + C_{23}d_4 - a_3S_{23} - a_2S_{23} - a_2S_2) \\ S_1(d_6(-S_{23}C_4S_5 + C_{23}C_5) + C_{23}d_4 - a_3C_{23}) \\ -d_6S_1C_{23}S_4C_5 + C_1d_6C_4S_5 \\ S_1(d_6C_{23}C_4C_5 - S_{23}S_5) + C_1d_6S_4C_5 \end{bmatrix}$$

$$Jv_z = \begin{bmatrix} 0 \\ d_6(-S_{23}C_5 + C_{23}C_4C_5) - S_{23}d_4 - a_3C_{23} - a_2C_2 \\ d_6(-S_{23}C_5 + C_{23}C_4C_5) - S_{23}d_4 - a_3C_{23} \\ -d_6S_{23}S_4S_5 \\ d_6(-C_{23}S_5 + S_{23}C_4C_5) \end{bmatrix}$$

Finally, substituting (12) with evaluated expressions we can identify the required mass matrices in operation space for given robot configurations. The mass matrix numerical representation with obtained eigenvalues is presented in (38):

$$M_v = \begin{bmatrix} 7.7 & -0.2 & -1.8 \\ -0.2 & 30 & -0.3 \\ -1.8 & -0.3 & 39 \end{bmatrix} \quad (38)$$

$$\lambda_1=40, \lambda_2=30, \lambda_3=7.6$$

$$\frac{1}{40} = [0 \ 0 \ 1] M_v^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (39)$$

It is seen that the maximum effective mass in the direction of the eigenvector $V(\lambda_1) = [0,0,1]^T$ (39) (Fig. 8 (a)) can not cause serious injury to a human (sphere H), however, if the configuration/direction is changed, as it is shown in Fig. 8 (b) ($V(\lambda_1) = [1,..,0]^T$), with no variations in the maximum effective mass value, personnel can be under a great risk to be injured.

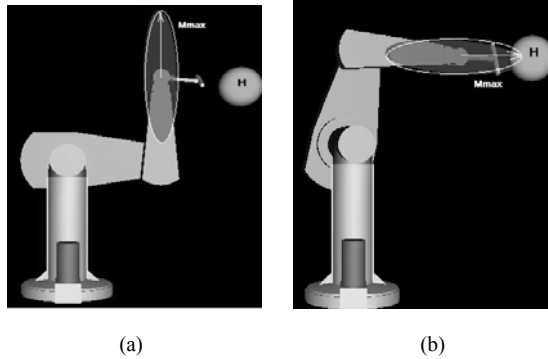


Figure 8

PUMA robot configurations: a) "safe motion/ configuration," b) dangerous motion

To estimate the level of this risk we refer to the acceleration and force related danger indices computation assuming that the manipulator interface stiffness K is 250 KN/m and the mass m of a human head is 5 kg.

7.1 Acceleration Related Danger Index Computation

From the simulation results it was obtained that the mean duration of an impact Δt is 0.025 s. Thus, the critical acceleration value was recomputed with respect to this value in (44):

$$a_c = 9.81 \times 2.5 \sqrt{\frac{265}{0.025}} < 390 \text{ m/s}^2 = 39g \quad (40)$$

Furthermore, basing on acceleration related danger indices evaluated in (26), (32) and considering that the maximum robot effective mass M is 40 kg, boundaries for an initial robot speed were defined:

$$\frac{M}{m} v_0 \omega \sqrt{\frac{M}{M+m}} \leq a_c; \quad v_0 \leq 390 \sqrt{\frac{M+m}{M}} \frac{m}{M\omega} \approx 0.7 \text{ m/s} \quad (41)$$

$$2v_0 \frac{M}{m} \sqrt{\frac{M}{m+M}} \leq a_c \Delta t; \quad v_0 \leq 390 \times 0.025 \sqrt{\frac{M+m}{M}} \frac{m}{2M} \approx 0.7 \text{ m/s} \quad (42)$$

From the graphical representations below it can be noticed that in spite of the equivalency of two definitions, there is a significant characteristics alteration in the condition when the critical level is overrun. In the HIC based formulation an extent of danger increases much greater. (See Fig. 9 (b))

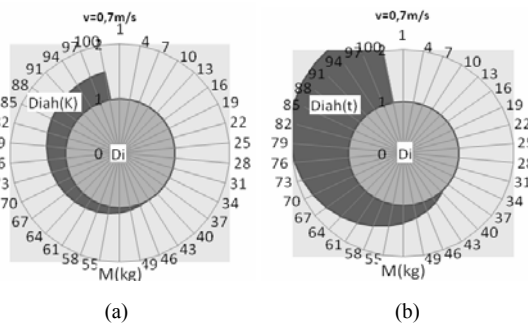


Figure 9

Acceleration related danger Index based on: a) collision modeling-based, b) HIC criteria-based

The impact force that corresponds to the estimated critical acceleration is estimated as 2 KN. This force is more than enough to cause fracture to the human facial bone. In the case of clamping (trapping) the extent of a penetration (σ) (computed according to (43)), with the facial bone stiffness $K_{fb}=100$ KN/m, can reach 20 mm, that is above the tolerable level.

$$\sigma = F / K_{fb} \quad (43)$$

Therefore, this criterion cannot be used under certain critical conditions. To meet more restrictive safety requirements, where no bone whatever stiffness it has can be under the risk to be fractured (or even no pain caused) a force related danger index should be applied.

7.2 Force Related Danger Index Computation

Computations were provided on the basis of evaluations presented in (33) for different critical forces including pain tolerance limits and the robot safety standardized requirements. In the (44) and (45) boundaries on the manipulator operating velocities were established based on the pain tolerance (130 N) and maxilla fracture limit forces (660 N) respectively:

$$v_o \leq 130 \sqrt{\frac{M+m}{M}} \frac{1}{M\omega_n} = 0,05 \text{ m/s} \quad (44)$$

$$v_o \leq 660 \sqrt{\frac{M+m}{M}} \frac{1}{M\omega_n} = 0,26 \text{ m/s} \quad (45)$$

Figure 10 illustrates the danger index behavioral characteristics according to identified boundaries assuming that the actual manipulator velocity is 0.7 m/s. In this case robot operating safe conditions can be only reached when an arm effective mass is not greater than 3 kg for a pain tolerance (See Fig. 10 (a)) and 14 kg for a maxilla fracture criteria (See Fig. 10 (b)).

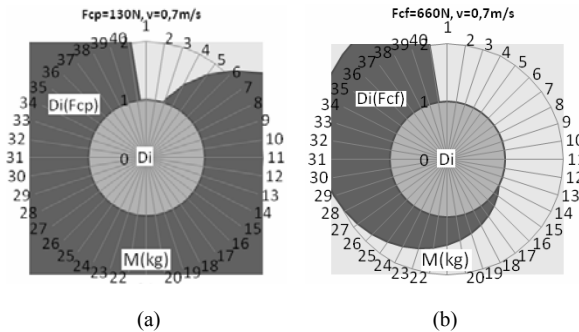


Figure 10

Force related danger index chart (case study) with critical forces $F=130\text{N}$ (a), and $F=660\text{N}$ (b)

With a reference to a safety standard [26], where robot speed should not exceed 0,25 m/s and exerted force -150 N, we can identify that only the force related danger index approach meets requirements of this standard. In the Fig. 11 four indices including standard requirements are represented. The mapping was provided for the robot speeds 0.14, 0.25, 0.6 and 1 m/s. Velocities 0.14 and

0.6 m/s were related to a psychological factor. Experimental researches from different groups [27] showed that at these velocities person does not feel fear or discomfort during interactions with robots. The speed 0.14 m/s was also associated with the first level of interaction (L1, collaborative), while 0.6 m/s was considered for a second level (L2, interior monitoring). From the charts in Fig. 11 it can be seen that at the velocities 1m/s all danger indices exceed the admissible level for the effective mass $M=40$ kg with interface stiffness $K=200$ KN/m. (See Tab. 2) At 0.6 m/s, only the acceleration related index had positive results. Meanwhile, slow end effector motions were found acceptable for all safety requirements.

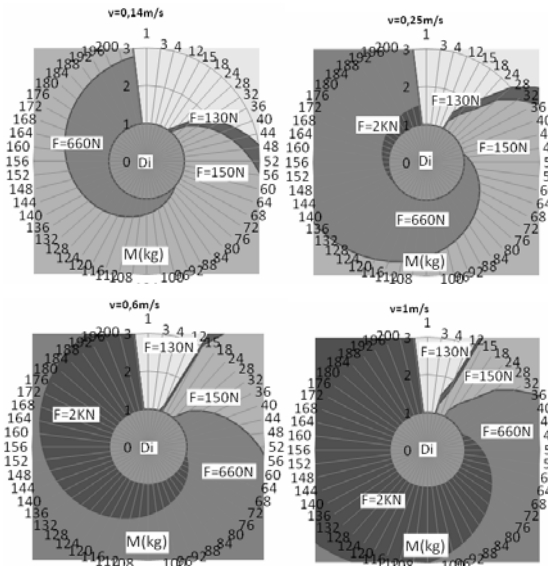


Figure 11

Danger indices comparison characteristics: force related (F (130,660 N)), ANSI/RIA standard (F(150 N)) acceleration related ($a_h=39$ g) for chart evaluated for a speed range: 0.14, 0.25, 0.6, 1 m/s

Table 2
Danger Indices Comparative Characteristic

Danger Index	Fc, N	Me(kg)							
		0,14m/s		0,25m/s		0,6m/s		1m/s	
Di(F_{cp})	130	15	-	8	-	3	-	2	-
Standard	150	18	-	10	-	4	-	2, 5	-
Di(F_{cf})	660	72	+	40	+	18	-	10	-
Di(F_{ah})	1950	238	+	136	+	56	+	33	-

'+' indicates the fulfillment of the danger criteria conditions for the manipulator effective mass 40 kg

The results showed that the only force related danger index approach meets requirements stated in the robot safety standard. If it is necessary for a task performance to increase robot speeds, with configuration where robot effective mass is relatively high, then to avoid the risk of serious injury an interface stiffness of this robot has to be lowered. For instance, if we apply a soft rubber material for a robot wrist with the stiffness 100 KN/m, speeds can be raised to the value up to 1 m/s as it is shown in the Fig. 12a ($M=40$ kg). From the diagram (Fig. 11b) it can be noticed that in this approach a head acceleration is reduced in almost 1,5 times.

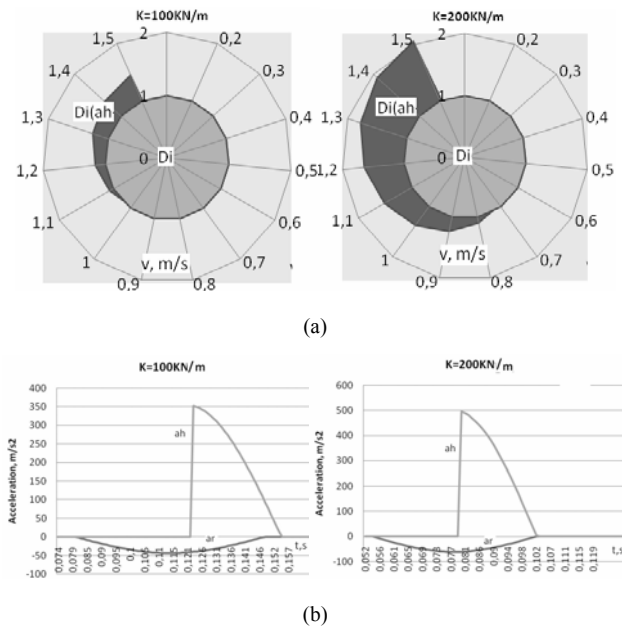


Figure 12

Effect of the Stiffness change on the admissible velocity range (a) and acceleration after impact (b)

8 Proposed Applications

To keep a “safe” level of interaction on the basis of provided estimations various strategies can be introduced. For instance, for large effective mass configurations, when the danger index is exceeded already at a relatively low velocity, and if there is a possible contact with a human, manipulator trajectory (points) should be redesigned to maintain a tolerable level of danger. Thus, the whole robot path (or, if it is hard to provide, near points) must be hazard free. Fig. 13 (a) illustrates a situation when the manipulator is moving in the direction where the human

presence is not acceptable (danger index circle is violated). In this case, when personnel are detected in the zone, robot control should whether stop the operation or make all possible corrections for the danger reduction. Fig. 13 b illustrates a field where human is allowed to approach a robot operating space. This area is resulted from an intersection of a danger index sphere and an effective mass ellipsoid, represented as a 3-D conic space with an angle φ . For instance, taking danger index associated with a critical pain tolerance force ($F_{cp}=130$ N) where permissive effective mass value is estimated as 8 kg, at a maximum robot operating speed 0.25 m/s along the direction u an angle φ will be 132 grad (See Fig. 13 b, c). However, zone outside this area should be restricted to prevent any non authorized entering. This approach does not require any on-line changes in the robot configurations or speed during task performance while personnel are inside the “safe” space. Fig. 14 displays a manipulator that is tracking a linear path (from A to B) with low effective mass (inertia) control and constant operating velocity (0.5 m/s here). During this motion a maximum effective mass is changing from 40 to 110 kg, however, effective mass at the direction of the following trajectory m_u , is controlled to not exceed the threshold value (8 kg for interaction Level 1). Thereby, this trajectory is “safe” from any harmful impact to a human, even if there is an unexpected robot motion takes place.

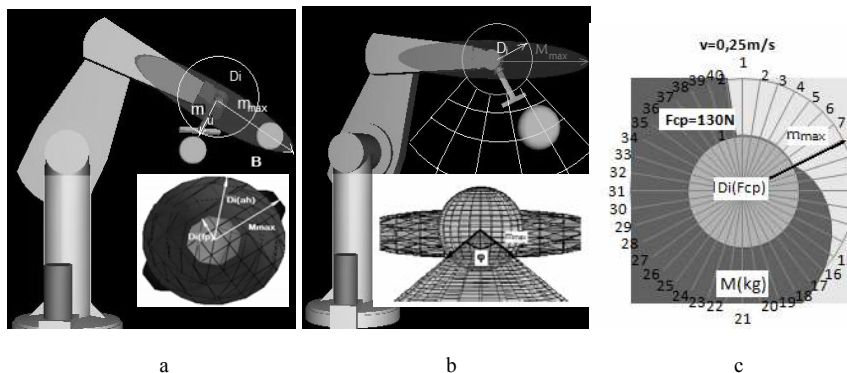


Figure 13

Robot safeguarding strategies: a) safety violation b) 3D conic field, c) danger index representation for pain tolerance criteria ($F_{cp}=130$ N)

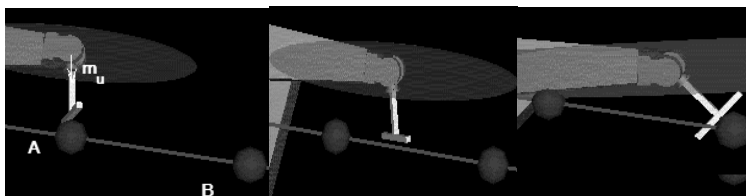


Figure 14

Robot “safe” path tracking

Conclusion

In the presented work three danger indices were developed and investigated. It was identified that even for very low or minor injury severity levels according to AIS scale, there is a risk to be injured at some critical conditions (trapping). Force related danger index is found to be more appropriate for these situations and closer to the robot safety standard requirements fulfillment. Introduced indices enable to provide analysis on robot operating hazardous characteristics and identify the extent of the potential task/robot related danger. Developed approach allows to human and robot collaborate within all interaction levels maintaining the risk and probability of an accident occurrence at a very low level.

In the future work it is planned to integrate this approach into a safety monitoring system, that would provide the faster and more reliable response of the robot system to the non anticipated failures and hazardous situations.

References

- [1] De Santis A., Siciliano B., De Luca A., Bicchi A.: Atlas of Physical Human-Robot Interaction, Mechanism and Machine Theory, Vol. 43, No. 3, March 2008, pp. 253-270
- [2] Albu-Schaffer A., Bicchi A., Boccadamo G., Chatila R., De Luca A., De Santis A., Giralt G., Hirzinger G., Lippiello V., Mattone R., Schiavi R., Siciliano B., Tonietti G., Villani L.: Physical Human-Robot Interaction in Anthropomorphic Domains: Safety and Dependability, 4th IARP/IEEE-EURON Workshop on Technical Challenges for Dependable Robots in Human Environments, Nagoya, July 2005
- [3] Zinn M., Khatib O., Roth B., Salisbury J.: Playing It Safe-Human-Friendly Robots, IEEE Robotics and Automation Magazine, Vol. 11, 2002, pp. 12-21
- [4] Carlsson J.: Robot Accidents in Sweden, National Board of Occupational Safety and Health, Sweden. Ed. M. Bonney, Yong Y. et al., Springer, Berlin, 1985, pp. 49-64
- [5] Hirschfeld R., Aghazadeh F., Chapleski R.: Survey of Robot Safety in Industry, International Journal of Human Factors in Manufacturing, Vol. 3, No. 4, March 2007, pp. 369-379
- [6] UAW Health and Safety Department: Review of Robot Injuries - One of the Best Kept Secrets, Proceed. National robot conference, Ypsilanti, Michigan, Oct. 2004
- [7] EuroNCAP: European Protocol New Assessment Programme-Assessment Protocol and Biomechanical Limits, 2003
- [8] Versace J.: A Review of the Severity Index, In Proceedings of the 15th Car Crash Conference, Society of Automotive Engineers, New York, 1971, pp. 771-796

-
- [9] Gurdjian E., Lissner H.: Mechanism of Head Injury as Studied by the Cathode Ray Oscilloscope, preliminary report, Journal of Neurosurgery, 1944, pp. 393-399
- [10] Gadd C.: Use of a Weighted Impulse Criterion for Estimating Injury Hazard, In Proceedings of the 10th Stapp Car Crash Conference, Society of Automotive Engineers, New York, 1966, pp. 164-174
- [11] McElhaney J., Stalnaker R., Roberts V.: Biomechanical Aspects of Head Injury, Human Impact Response - Measurement and Simulation, 1972
- [12] Prasad P., Mertz H.: The Position of the USA Delegation to the ISO Working Group on the Use of HIC in the Automotive Environment. Society of Automotive Engineers Technical, 851246, Warrendale, PA, 1985
- [13] AIS for the Advancement of Automotive medicine, The Abbreviated Injury Scale, Revision Update 1998, Des Plaines/IL
- [14] Haddadin S., Albu-Schäffer A., Hirzinger G.: The Role of the Robot Mass and Velocity in Physical Human-Robot Interaction – Part I: Non-constrained Blunt Impacts, in IEEE International Conference on Robotics and Automation ICRA, Pasadena, USA, 2008
- [15] Zinn M.: A New Actuation Approach for Human Friendly Robotic Manipulation, PhD thesis, Stanford University, CA, 2005
- [16] Ikuta K., Ishii H., Nokata M.: Safety Evaluation Method of Design and Control for Human-Care Robots, International Journal of Robotic Research, Vol. 22, No. 5, 2003, pp. 281-298
- [17] Yamada Y., Hirasawa Y., Huang S., Uematsu Y., Suita K.: Human-Robot Contact in the Safeguarding Space, IEEE/ASME Trans. on Mechatronics, Vol. 2, No. 4, 1997, pp. 230-236
- [18] McCormick E., Sanders M.: Human Factors in Engineering and Design, McCraw-Hill, 1992
- [19] Suita K., Yamada Y., Tsuchida N., Imai K., Ikeda H., Sugimoto N.: A Failure-to Safety 'Kyozon' System with Simple Contact Detection and Stop Capabilities for Safe Human Autonomous Robot Coexistence, IEEE Int. Conf. on Robotics and Automation, 1995
- [20] Hertz E.: A Note on the Head Injury Criterion (HIC) as a Predictor of the Risk of Skull Fracture, In 37th Annual Association for the Advancement of Automotive Medicine, Association for the Advancement of Automotive Medicine, 37, Plaines, IL, 1993, pp. 303-312
- [21] Haddadin S., Albu-Schäffer A., Hirzinger G.: The Role of the Robot Mass and Velocity in Physical Human-Robot Interaction-Part II: Constrained Blunt Impacts, IEEE Int. Conf. on Robotics and Automation ICRA, 2008

- [22] Haddadin S., Albu-Schäffer A., Strohmayer M., Hirzinger G.: Approaching Asimov's 1st Law II: The Impact of the Robot's Weight Class, IEEE Int. Conf. on Robotics and Automation ICRA, 2008
- [23] Khatib O.: Inertial Properties in Robotic Manipulation: an Object-Level Framework, Int. J. Robotics Research, Vol. 14, No. 1, 1995, pp. 19-36
- [24] Khatib O., A. Bowling: Optimization of the Inertial and Acceleration Characteristics of Non-Redundant Manipulators, Proc. 3rd Conference on Mechatronics and Robotics, Paderborn, Germany, Oct. 1995, pp. 500-510
- [25] Armstrong B., Khatib O., Burdick J.: The Explicit Dynamic Model and Inertial Parameters of the PUMA 560 Arm Robotics and Automation, Proc. of IEEE International Conference, Vol. 3, 1986, pp. 510-518
- [26] ISO10218, Robots for Industrial Environments - Safety Requirements - Part 1: Robot, 2006
- [27] Nagamachi M.: Human Engineering-oriented Research on Industrial Robots, Human Engineering, Vol. 19, No. 5, 1983, pp. 259-64

Mechanical Pretreatment of Surface of Aluminum Alloy D16-T by Shot Peening

Daniel Kottfer¹, Peter Mrva²

¹Department of Technologies and Materials
Faculty of Mechanical Engineering
Technical University of Košice
Mäsiarska 74, 040 01 Košice, Slovakia, e-mail: daniel.kottfer@tuke.sk

²Department of Aviation Engineering
Faculty of Aeronautics
Technical University of Košice
Rampova 7, 041 21 Košice, Slovakia, e-mail: peter.mrva@tuke.sk

Abstract: The paper describes the influence of shot-peening onto aluminium alloy D16-T surface. There are estimates focusing into microgeometry of the shot peened surfaces, their roughness and a size of selected shot peened material - the corund. Based on the results of measurements, the evaluation was oriented on the curve of roughness, functionality of the surface roughness R_a and the necessary quantity of shot peening material q_{nR} of estimated material depending on the grain size d_z .

Keywords: shot peening, functional surface, surface roughness

1 Introduction

Surface strengthening by shot peening can make use shot peening to increase resistance to fatigue stress of engineering components. Some alloys (on the basis of magnesium) are inclining to fatigue cracks. Defects like grains and systemless structures begin and accelerate this cracks – tension focusing.

By high frequentional cyclical density, for smallest numbers of cycles to crash locations of cracks start appearing on the surface [1, 2]. For higher numbers of cycles to crash are locations of cracks begin to appear in the thermal area of the experimental sample [3]. After thermal treatment of some alloys the obtained structure, the compound of which is balanced polyedric grains with concrete phases. It involves mechanical properties growth and resistance to fatigue as well [4].

In present time there is enough developed imagination of deposition process of thermal spraying coatings on the surfaces of steel engineering accessories.

Lifetime of coatings depend on the ideal adhesion of functional coatings [5, 6, 7]. Adhesion is conditioned by ideal pretreatment of the functional surface. Shot peening is one of the most frequently used technologies of mechanical pretreatment of surface under thermal spraying coatings. The surface is cleaned by shot peening. It is created applicable microgeometry of the surface, too. It is known that the activation energy of surface made from deformation of surfaces layers during shot peening is definitely influencing on the coating adhesion [8]. This energy value reduces exponentially as a result of background influence. Therefore coating is to be deposited between 1-3 hours. Adhesion of coating to basic material can be evaluated by mechanism of adhesion. Mechanical adhesion of coating in surface relief of sample takes 50-80%. The Van der Waals forces make about 5% and power of chemical compounds make up 15-45%. In comparison with surface pretreatment by cauterization, shot peening is more convenient than cauterization. The technology of surface shot peening can be used for surface strengthening and producing good roughness of surface. The experiment was focused on the research of the microgeometry of surface, character of surface, influence of concrete parameter for making good, strong coating and the substance.

To use of thermal spraying technologies are actual for aviation components renovation, made from light alloys on the base Al, Mg, Ti. Experiment research the influence of sorts and dimensions shot peening material grain and shot peening parameters on the necessary quality of surface under thermal spraying coatings.

2 Experiment Methodology

Experimental research was headed to analyse shot peened coating microgeometry and the influence of technological parameters on the required quality of surface.

2.1 Evaluation of Shot Peened Coatings Microgeometry

Shot peening is a specific form of coatings pretreatment of components. Character of shot peened surface is typical for this technology. In the shot peening process the component surface is hacked. Roughness is evaluated by a touch-profiler. These appliances have bigger scale of measured parameters of surface roughness, for example $R_a=30\ \mu\text{m}$. Middle arithmetic aberrance R_a was selected for surface evaluation. For measuring values and for making profigrams of shot peened surfaces, profiler HOMMEL Tester T3 was used.

To obtain relevant results the next conditions have been used [7, 8]:

- length of measured distance $L = 6,3$ mm,
- terminal undulation (cut-off) $l = 1,25$ mm
- number of measurements $n = 10$.

Poligrams were taken out following next conditions:

- length of measured distance $L = 6,3$ mm,
- terminal undulation pinch $l = \infty$.

Medium arithmetic value as a statistic value has been calculated from measuring values of roughness. Each surface was evaluated by two profilegrams.

2.2 Experimental Samples Preparation

Dural samples D16-T (STN EN 42 49 11) were used in the experiment. Sample dimensions were chosen in such way to eliminate unwanted influence of shot peening device in process of shot peening (e.g heterogeneous consistency of the grain touches in the entire field of shot peening beam). Samples dimensions enabled as to realize adhesion test after thermal spraying coatings on surfaces of shot peened samples. Samples were made by turning into the form of a roll with diameter of 30 mm (Fig. 1). The functional surface of the samples before shot peening had the roughness of $R_a = 0,6$ μm .



Figure 1

Sample of the D16-T aluminium alloy after shot peening

2.3 The Material Used for Shot Peening

There are not uniform selected criteria for shot peening material by now. Selection of the shot peening material was based on the basic material properties.

For surfaces' shot peening process a corund granular was used (STN EN 22 40 12). This material is produced in all granularities and the shot peening material is a polydisperse. To explain the influence of the grain size into the surface roughness was the shot peening material selected by wire screen with a specific grain diameter. The chosen grain diameter was according to STN 15 3105.

2.4 Shot Peening Process

For shot peening of the evaluated sample surfaces a laboratory equipment [5] was used. The influence of the sort of the shot peening material was tested during the experiment at the following speeds: $v_1=78,1 \text{ ms}^{-1}$, $v_2=95,5 \text{ ms}^{-1}$ a $v_3=112,4 \text{ ms}^{-1}$. Grain angle incidence of shot peening material on sample surface was $\alpha=75^\circ$. The sample distance to the shot peening wheel was $L=200 \text{ mm}$. By Matling and Steffens, one cannot prevent hobbing of the shot peening material onto the shot peened surface. It can make galvanic cells [9]. Impresses (by hobbing) can be made by harder shot peening materials, too. Therefore for shot peening of the D16-T material corund was used with a minimum grain speed of $v_1=78,1 \text{ ms}^{-1}$. Next, corund of medium grain diameters: 0,36; 0,56; 0,71; 0,9; 1,12 mm were used.

3 Methodics of Determining the Hacking Curve of the Shot Peened Surface

The experiment was aimed to determination of the necessary quantity q_{HR} of abrasive material, which is needed to completely cover the shot peened surface. Initiate account was determined from character hackingcurves, with completing by visual scan by Meopta stereomicroscope with zoom 100x. The Hacking curve technique specifies the functional dependancy of the shot peened surface roughness onto the quantity of the abrasive material, which shapes the measured surface (Fig. 2). Area from first to second part is important for the determination of covering surface shot peening grade by Hackingcurves (Fig. 2) [6].

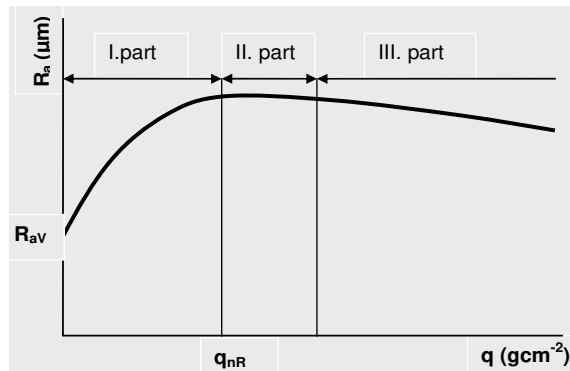


Figure 2
The hacking curve

3.1 D16-T Material Hacking Curves Determination

Samples from the D16-T material were shot peened gradually by amount of 1000 g, with covering grade $q=0,5 \text{ gcm}^{-2}$, number of amount 10. Next samples were shot peened by two amounts of 1000 g, fraction diameter $d_z=0,36; 0,56; 0,71; 0,9$ and $1,12 \text{ mm}$. After each shot peening the roughness R_a was measured. Each of R_a is the arithmetic average of 10 measured accounts. After each shot peening the surface was evaluated by means of optical microscope. After each shot peening another material was selected. The number of amounts in the experiment was selected so that the hacking curve could capture the first and second part, and partially the third one (Fig. 2). Hacking curves are in Fig. 3.

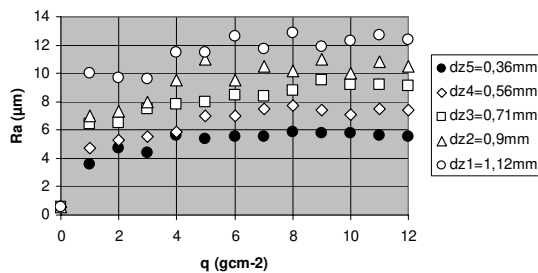


Figure 3
D16-T hacking curves with corundum fractions of diameters d_z

3.2 Necessary Quantity of an Abrasive Material Determination

To cover of shot peened surface with touches by shot peening, it is necessary to know the microgeometry of the shot peened surface. The grade of covering of this surface is to be $n=1$. Now, the necessary quantity of abrasive material for surface covering is on the hackingkurve q_{nR} . It is expected that the linear and planar covering grade is 1 (Fig. 2).

The dependency of determining the necessary quantity of the abrasive material in terms of the grain dimension can be solved from the hackingkurves (Fig. 3). The necessary quantity of abrasive material q_{nR} for dimension of the grain tested d_z was determined (Fig. 4). This way is valid for the shot peening applied.

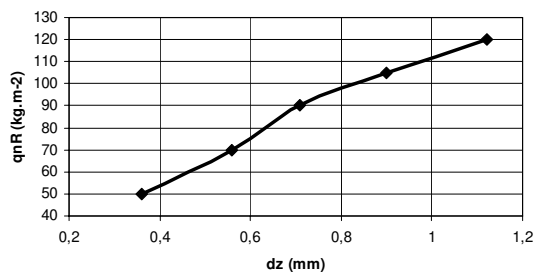


Figure 4

The correlation of necessary quantity of shot peening material q_{nR} on grain diameter d_z

3.3 Grain Diameter Influence of the Shot Peened Material on the Roughness R_a

The roughness value R_a of shot peened surface was dedicated as an arithmetic average from 10 measurements. Finished values for concrete grain diameters of shot peening material by speed $v_1=78,1 \text{ ms}^{-1}$ are in diagram (Fig. 3). There are necessary quantities involved of shot peening material q_{nR} , too. From function dependency it follows that as the grain diameter grows than the roughness value of shot peened material growth too. It is related directly with the touch size after the grains of the shot peening material falls on the surface.

4 Experimental Results Discussion

In the shot peening process were hacked samples surfaces gradually. So one can state that roughness change is different in measured values (Fig. 4). That were given material properties of the evaluated samples. The values R_a , q_{nR} , and d_z are in Table 1.

Table 1
 d_z , R_a and q_{nR} values of the material D16-T

d_z [mm]	0,36	0,56	0,71	0,9	1,12
R_a [μm]	5,8	7,5	9,25	10,8	12,5
q_{nR} [kgm^{-2}]	50	70	90	105	120

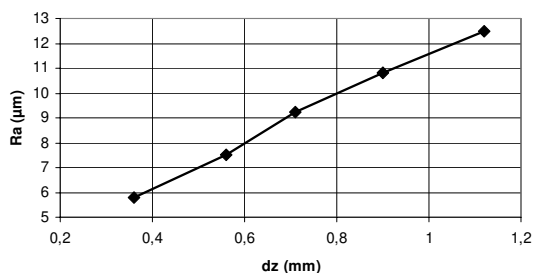


Figure 5

The correlation of surface roughness R_a of D16-T material on grain diameter d_z .

Conclusions

Upon analyses, study and realized experiments, can be said:

- the measured accounts were the basis for drawing the hacking curves – the change depends on the roughness R_a on the quantity of the shot peening material q_{nR} , which fall on the measured surface,
- the change in the roughness of the shot peened surface is influenced by the size of the grain of the shot peening material, as the size of diameter grows so does the roughness of the shot peened surface,
- necessary quantity of shot peening material q_{nR} for full covering of shot peened surface can be determined by the hacking curves,
- necessary quantity of shot peening material q_{nR} (corund) influences: shot peening parameters, especially the grain diameter,
- for the shot peened material D16-T a corund is the most suitable substance. By it can be achieved the cleanness of the shot peened surface and less necessary quantity of abrasive material q_{nR} with high accounts of roughness.

Acknowledgement

The article has been developed within the framework of solving tasks under AV No. 4/2021/08.

References

- [1] Piľa J., Sloboda A., Sloboda A.: Some Opportunities of Diagnostic Parameters Utilization in the Aircraft Proactive Maintenance Management. In: DIS 2004 : Teória a aplikácia metód technickej diagnostiky: 7. ročník medzinárodnej vedeckej konferencie, Košice, October 13-14, 2004. Košice: Dom techniky ZSVTS, 2004, pp. 76-82, ISBN 80-232-0237-5
- [2] Neštrák D., Piľa J.: Helicopter Aerodynamics, Structures and Systems (in Slovak): Textbook: Akademické nakladateľství CERM, 2006, p. 454, ISBN 80-7204-484-2
- [3] Kuffová, M., Bella, V., Wolny, S.: Fatigue Resistance of Mg–Alloy AZ 63HP under High-Frequency Cyclic Loading. In *Mechanika Kwartalnik Akademii Górniczo-Hutniczej imienia Stanisława Staśika w Krakowie*, 23, 3, 2004, Poland
- [4] Kuffová, M.: Microstructure of Magnesium Alloys after Heat Treatment. In *proc. "Opatřebení, spolehlivost, diagnostika 2006*, Brno, p. 139, ISBN 80-7231, Czech Republic
- [5] Mrva P., Kaliský S.: Mathematical Model Evaluation of the Shot Peening Technological Process of Ti Alloys onto Adhesion of Plasma Sprayed Coatings with Thermal Insulating Properties (in Slovak), Corosion and Corrosion Protection of Matrials (in Slovak), 4th International Conference, Trenčín, April 12-13, 2000, pp. 124-128
- [6] Kniewald D., Pivoda P.: Mechanical Pretreatment of the Surface of Parabolic Springs under Protective Al and Zn Thermal Sprayed Coatings (in Slovak), In.: *Zborník vedeckých prác VŠT v Košiciach*, 1978, pp. 309-318, Slovak Republic
- [7] Mrva.: Research of the Influence of the Surface Pretreatment of Titanium Alloys (in Czech), Research report VU 070 Brno, 1986, Czech Republic
- [8] Kniewald D., Šefara M.: Vorbehandlung der Metalloberfläche durch Strahlen als Vorbereitung für Schutzüberzüge aus Pulverkunststoffen, *Zborník vedeckých prác VŠT v Košiciach*, 1980, pp. 263-274, Slovak Republic
- [9] Sedláček V.: *Metall Surfaces and Coatings* (in Czech), ČVUT Praha 1992, ISBN 1335-2393

Customer Relationship Management: Implementation Process Perspective

Alok Mishra, Deepti Mishra

Department of Computer Engineering
Atilim University, Ankara, Turkey
alok@atilim.edu.tr, deepti@atilim.edu.tr

Abstract: Customer relationship management (CRM) can help organizations manage customer interactions more effectively to maintain competitiveness in the present economy. As more and more organizations realize the significance of becoming customer-centric in today's competitive era, they adopted CRM as a core business strategy and invested heavily. CRM, an integration of information technology and relationship marketing, provides the infrastructure that facilitates long-term relationship building with customers at an enterprise-wide level. Successful CRM implementation is a complex, expensive and rarely technical projects. This paper presents the successful implementation of CRM from process perspective in a trans-national organization with operations in different segments. This study will aid in understanding transition, constraints and the implementation process of CRM in such organizations.

Keywords: Customer Relationship Management, Customer, CRM, Implementation

1 Introduction

Companies that enter to compete in a new market weaken the existing and solid ones, due to new ways of doing and conceiving businesses. One of the factors that have driven all these changes is the constant change and evolution of technology. Because of this reality, the CRM concept has evolved in such a way that nowadays it must be viewed as a strategy to maintain a long-term relationship with the customers [1]. A good customer relationship is the key to business success. Relationship building and management, or what has been labelled as relationship marketing, is a leading approach to marketing [2]. The use of customer relationship management (CRM) systems is becoming increasingly important to improve customer life time value [3]. Understanding the needs of customers and offering value-added services are recognized as factors that determine the success or failure of companies [4]. So more and more businesses begin to attach great importance to electronic customer relationship management (eCRM), which focuses on customers instead of products or services, that is,

considering customer's needs in all aspects of a business, ensuring customers' satisfaction. By providing information on customer data, profiles and history they support important areas of a company's core processes, especially in marketing, sales and service [5]. eCRM is all about optimising profitability and enabled businesses to keep customers under control, as it makes the customer feel they are really a part of the business progress [6]. When managing the transition to a customer-centric organization, it is mandatory to develop the capabilities to acquire the necessary resources, knowledge and tools to meet customer's requirements with the appropriate products and services [1]. A knowledge based system is most effective in the managing of semi-structured problems. The abilities of such systems are usually applied on the managing level of strategic planning [7]. An effective CRM system should enable an organization to gain greater insight into customer behaviour and preferences whereas ERP analytics are more likely to focus on supply and demand for key resources and materials [4].

In spite of the wide use of sales force automation systems in sales [8], a Forrester study [9] observes significant deficits in today's marketing, sales and service processes. It was found that just 22% of the companies surveyed possess a uniform customer view and only 37% know which customers are looked after by individual business units [10]. To eliminate weaknesses in customer contact, many companies are either planning or in the process of implementing CRM systems. According to Gartner survey [11], 65% of US companies intended to initiate CRM projects in 2002. In Europe, roughly 3% of companies had fully implemented a CRM project in 2001, 17% had initiated more than one local project and 35% were developing concepts for the introduction of CRM [12]. The software CRM market is expected to increase from \$7 billion in 2000 to 23 billion in 2005, even though conventional wisdom is that 30 to 50 percent of CRM initiatives fall short of meeting company objectives, while another 20 percent actually damage customer relationships [13].

Different organizations are approaching CRM in different ways. Some view CRM as a technology tool while others view it as an essential part of business. According to Verhoef *et al.* [14], the success rate of CRM implementation varies between 30% and 70%. According to industry analysts, almost two-thirds of CRM system development projects fail [15]. According to IDC (International Data Corporation) and Gartner Group, the rate of successful CRM implementations is below 30% [16], hardly justifying the cost of implementation [17]. Another report estimates that between 60 and 90 percent of enterprise resource planning implementations do not achieve the goals set forth in the project approval phase [18] Hence, key factors of success or failures during CRM implementation have been the subject of active research in recent years [19]. The study performed by Forsyth took a sample of 700 companies, with regards to the causes of failure to reach the CRM benefits [20]. The main causes of failure were:

- Organizational change (29%)

- Company policies/inertia (22%)
- Little understanding of CRM (20%)
- Poor CRM skills (6%)

The results show that there is no ‘unique’ CRM project and that successful implementations are rarely technical projects [10]. Therefore the objective of this paper is to report successful CRM implementation and lessons learned in an organization involved in many countries with operations in different segments.

CRM is a synthesis of many existing principles from relationship marketing [21], [22], [23] and the broader issue of customer-focused management. CRM systems provide the infrastructure that facilitates long-term relationship building with customers. Some examples of the functionality of CRM systems are sales force automation, data warehousing, data mining, decision support, and reporting tools [24], [25]. CRM systems also reduce duplication in data entry and maintenance by providing a centralized firm-database of customer information. This database replaces systems maintained by individual sales people, institutionalizes customer relationships, and prevents the loss of organizational customer knowledge when sales people leave the firm [26]. Centralized customer data are also valuable to firms managing multiple product lines. In many cases customers will overlap across different lines of business, providing an opportunity for increasing revenues through cross-selling.

The paper is organized as follows: Section 2 reviews the literature on CRM implementation. In Section 3 we have presented the CRM implementation in a multinational organization. Finally Section 4 draws conclusions from the case study in terms of its practical relevance and lessons learned.

2 Literature Review

The first requirement for the successful implementation of CRM is clarity regarding CRM terminology. From the many approaches available, the distinction between the following three areas has become generally accepted [27].

- **Operational CRM** supports front office processes, e.g. the staff in a call center. Operational integration points exist to human resource systems for user data and ERP systems for transferring order information which was captured e.g. from a call center representative [10]. From an operations perspective, Bose [28] pointed out that CRM is an integration of technologies and business processes that are adopted to satisfy the needs of a customer during any given interaction.

- **Analytical CRM** builds on operational CRM and establishes information on customer segments, behaviour and value using statistical methods. It is useful for management and evaluation purposes, the operational customer data are integrated with a centralized data warehouse which is consolidated data based on certain criteria (e.g. sales, profits). Here the data mining tool analyses defined dimensions, e.g. compares the characteristics of one customer with another, leading to the determination of a customer segment and thus providing the basis for a targeted marketing campaigns [10].
- **Collaborative CRM** concentrates on customer integration using a coordinated mix of interaction channels (multi-channel management), e.g. online shops, and call centres. Approximately 60% of the companies surveyed use internet portals in their customer communication for selected or suitable activities [10].

CRM is therefore understood as a customer-oriented management approach where information systems provide information to support operational, analytical and collaborative CRM processes and thus contribute to customer profitability and retention. While potential benefits are attractive, CRM implementation must be managed carefully to deliver results [4].

Automation refers to using technologies including computer processing to make decisions and implement programmed decision processes [29]. The CRM system is the automation of horizontally integrated business processes involving “front office” customer touch points –sales (contact management, product configuration), marketing (campaign management, telemarketing), and customer service (call center, field service)-via multiple, interconnected delivery channels. Therefore, CRM system implementation is commonly used in functional areas such as customer support and services, sales and marketing. CRM life cycle includes three stages: Integration, Analysis and Action [30]. In the first stage, The CRM lifecycle begins with the integration of front office systems and the centralization of customer-related data [19]. Second stage called Analysis is the most critical to CRM success [30]. CRM analytics enable the effective management of customer relationships [19]. Using CRM analytics, organizations are able to analyse customer behaviours, identify customer-buying patterns and discover casual relationships [30]. The final phase, Action, is where the strategic decisions are carried out. Business processes and organizational structures are refined based on the improved customer understanding gained through analysis [31]. This stage closes the CRM loop and allows organizations to cash in on the valuable insights gained through analysis. Systemic approaches to CRM help organizations coordinate and effectively maintain the growth of different customer contact points or communication channels. The systemic approach places CRM at the core of the organization, with customer-oriented business processes and the integration of CRM systems [32].

According to Gefen and Ridings [33], a CRM system consists of multiple modules including: operational CRM, which supports a variety of customer-oriented business processes in marketing, sales and service operations; and analytic CRM which analyses customer data and transaction patterns to improve customer relationships. Operational and analytic CRM modules provide the major functions of a CRM system. Successful CRM implementation often entails significant organizational transformation due to the complexity of multiple operations involved in managing customer relationships [34]. Implementing a CRM system is only part of the needed change. To adopt the new ways of interacting with customers, firms need to align various organizational aspects with their CRM systems, e.g. business processes, strategies, top management support, and employee training [35]. A typical CRM implementation can be classified into six iterative processes including exploring and analysing, visioning, building business case, planning and designing solution, implementing and integrating, and realizing value [31]. Resulting from a variety of catastrophic ERP implementation failures, research on ERP systems points to the need to reduce application complexity. The likelihood of success is related to reduced project scope, complexity, and customization of the application. Defining a reasonable (i.e., smaller) system scope by phasing in software functionality over a series of sequential implementation phases is an important means of decreasing complexity. Similarly, reducing or eliminating customization of the specific functionality of CRM application software is critical to lowering risk. It is business needs that should determine the CRM application functionality – the scope of functions to be implemented [36]. Organizations are finding that implementing CRM functionality beginning with quick, clear-cut and profitable ‘hits’ helps to insure the initial success, and thus long- term success of a CRM initiative.

Generally, the case study method is a preferred strategy when “how” and “why” questions are being posed, and the researcher has little control over events [37]. The case study method, a qualitative and descriptive research method, looks intensely at an individual or small participants, drawing conclusions only about the participants or group and only in the specific context [37]. The case study method is an ideal methodology when a holistic, in-depth investigation is required [38]. Case studies are often conducted in order to gain a rich understanding of a phenomenon and, in information systems research, the intensive nature, the richness of a case study description and the complexity of the phenomenon are frequently stressed in case study reports [39].

3 Case Study

3.1 Organization Background

Organization is a trans-national enterprise with operations in different segments. This company engages in the design, manufacture, and sale of precision motion and fluid controls, and control systems for various applications in markets worldwide. The company has been growing rapidly in all segments.

3.2 Information Technology Infrastructure

The company has highly skilled engineers and has grown from being a small to a large company. IT in the company has been “home-grown”, i.e. systems were created using available tools to capture processes. Employee empowerment is very high in the company. This also meant that the company’s business units could decide what systems – hardware, software and networks it wanted individually. This has led to a plethora of IT systems. The CIO (Chief Information Officer) of the company started rationalizing the “basic” infrastructure to Lotus Notes for e-mail and Microsoft Office Suite for office applications. An Enterprise Resource Planning System (ERP) from QAD called MFG/PRO was implemented to take care of manufacturing, financials and logistic transactions of the company. This system was implemented individually in each country. Customization was not allowed without confirmation by a change request committee. Since reporting in MFG/PRO was weak, the company went ahead with a data-warehousing solution called “Cubes” based on a Progress database. Data needed for financial and management reporting was extracted on a daily basis from MFG/PRO into the cubes for analysis. The group is now considering moving all the disparate MFG/PRO systems to its data centre in the main office.

3.3 The Search for IT Solution

The company has doubled its’ operations over the past five years. The growing number of customers in various segments calls for a solution in information technology (IT). The company has over 5,000 customers spanning various markets like Power, Plastics, Metal Forming, etc. Due to large number of customers using the company’s components in various markets for various applications and lesser profitability, it was decided to bring together senior managers in the company for determining its future strategy for the IT solution. They found that use of IT in CRM would help the company in maximizing revenues in a cost-effective manner through various applications to a consolidated database, for example, sales forecasting, decision of marketing strategies, and customer identification.

3.4 Impetus for CRM

CRM can be defined as a management process of acquiring customers by understanding their requirements; retaining customers by fulfilling requirements more than their expectations; and attracting new customers through customer specific strategic marketing approaches. This requires total commitment from the entire organization. CRM uses IT to track the ways in which a company interacts with its customers; analyses these interactions to maximize the lifetime value of customers while maximizing customer satisfaction. The company has a large customer base, though the value of business from each customer is currently low. CRM would help the company in identifying customers who provide the greatest revenues for every marketing or service dollar spent or customers who cost little to attract. Typically, these 'good' customers present 80 to 90 percent of the company's profits, though they are only 10 to 20 percent of the client base.

The motivation for selecting CRM in the company was to increase business value due to the following:

- Information about customers is stored in disparate applications as the employee empowerment is very high. This customer related information from various systems needed to be brought in, analysed, cleansed and distributed to various customer touch-points across the enterprise, so that the various stakeholders – marketing, sales and engineering teams see a single version of 'truth' about the customer.
- This single source of customer data can be used for sales, customer service, marketing, etc. thereby enhancing customer experience and reducing churn-rate. Churn-rate measures the number of customers who have stopped using the company's products.
- By storing information about past purchases, sales team can make customized selling or personal recommendations to the customer. Also, this helps in up-selling or cross-selling opportunities.
- Capability to improve current sales forecasting, team selling, standardizing sales and marketing processes and systems.
- Support direct-marketing campaigns by capturing prospect and customer data, provides product information, qualified leads for marketing, and scheduling and tracking direct marketing communication. Also, it helps the marketing team fine-tune their campaigns by understanding the prospect of customer conversion.
- To help engineering in understanding market demand for specific product designs and act accordingly.
- Single out profitable customers for preferential treatment, thereby increasing customer loyalty.
- Easing sales account management through consolidated information.

3.5 CRM Implementation Process

3.5.1 ERP Selection

Since there were two different ERP systems in the company, with one mail system, it was difficult for the company to choose the right CRM system. In the end, a relatively unknown system called Relavis was selected as the preferred ERP system. Relavis was chosen because it tightly integrated with IBM Lotus Notes which is the common infrastructure across the whole enterprise. Relavis is a small company. The product is more economical than a Seibel, SAP or Oracle. The system has modules to cater to eMarketing, eSales and eService.

3.5.2 Scoping

The scope covered sales and marketing processes and followed the ‘**service platform**’ approach. A service platform integrates multiple applications from multiple business functions (in this case, sales, marketing, engineering), business units or business partners to deliver a seamless experience for the customer, employee, manager or partner. As shown in Figure 1, the new system (Relavis) was implemented to gain integrated information from marketing and sales departments to provide input to the ERP and Data warehousing applications and finally create analytical reports to make better business decisions e.g. to understand the sales results of specific leads, recommend better selling techniques and target specific leads etc. The new application could track the status of a lead through all stages of the sales and marketing lifecycle.

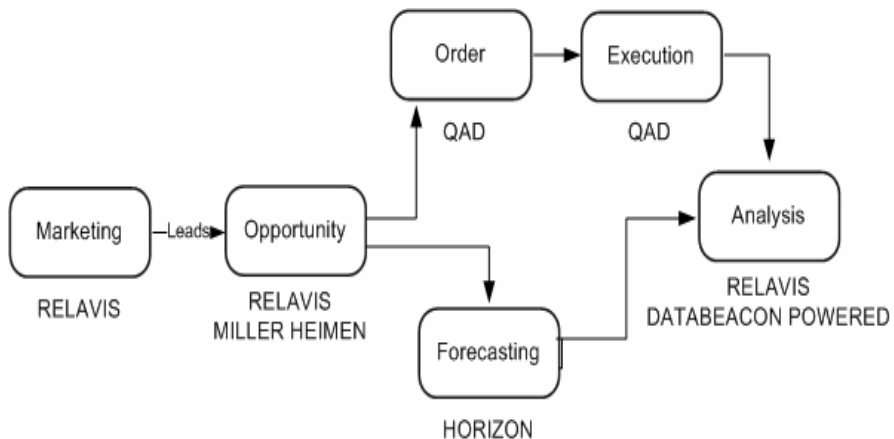


Figure 1

Enterprise's CRM Implementation Overall Process Flow [40]

Marketing was working on branding strategies and segmentation. Events were managed by marketing. These events would come up with a huge number of leads for new opportunities and marketing wanted to handover leads to sales. Sales filtered the leads from marketing and their own sources into opportunities. Opportunities were defined as those having specific sales persons assigned. These accounts were carefully evaluated to see if they fit with company's overall strategy of increasing revenue and profitability by solution selling. The Miller Heiman Process was used to capture relevant information on the opportunity and the blue-sheets of Miller Heiman were closely monitored by VP Sales and top management. The non-strategic product sale was channelled to distributors and agents. Consolidated forecast numbers were reviewed by senior management on a regular basis. Orders that were received were executed.

3.5.3 Design

A "gap analysis" was conducted since the CIO (chief information officer) wanted a successful "business" implementation of the system vis-à-vis a technical implementation, the sales and marketing process was mapped. The "as-is" process described the cradle-to-grave aspects of the process. The "to-be" process incorporated Relavis, together with other tools like Miller Heiman eforms, The Horizon system for forecast, MFG/PRO system for order execution and Datawarehouse Cubes for analysis. Relavis was customized to include "Business Intelligence" – a piece of software extracting account specific information from past sales through the Cubes.

3.5.4 Implementation

Implementation involved reviewing the resource requirements and availability, both in terms of hardware and software. The company had Lotus Notes skills in the organization. The system was simple. Hence the implementation was done using in-house resources. Training on the product was arranged from Relavis and its partners. The system approach involved a "big-bang" approach. After all, an audit and review should be undertaken to determine the monetary as well as non-monetary benefits against costs incurred. The implementation primarily consisted of the major steps as given in Table 1.

3.5.5 Impact

The system was packaged software, with very minimal customization. The only additions to the software were the Business Intelligence part and electronic Miller Heiman blue-sheet for strategic opportunities and gold-sheets for Large Accounts.

Some key users were involved in the decision-making. The project implementation plan was received well by all. The IT department made sure that the project was driven by sales for the eSales module and marketing for the

eMarketing module. A steering committee comprised of senior managers of each country (called REPCOTE or Relavis Pacific CORE TEam) was formed to drive the implementation. IT took the role of being facilitator.

Table 1
Major tasks during implementation and their duration [40]

ID	Task Name	Duration
1	Infrastructure readiness	2 days
2	Get Licenses	1 day
3	Synch with Lotus Notes team on training	0.1 days
4	Client PCs / Notebooks and network connecti	0.1 days
5	Give training sys implementn doc to local IT s	1 day
6	Process Mapping	5 days
7	Sales process data gathering	3 days
8	"To-be" Sales process mapping	1 day
9	"To-be" Support process map	1 day
10	Data cleanup	15 days
11	Send excel formats to countries	1 day
12	Identify account-supporting documents to be	1 day
13	Existing - MFG/PRO data	4 days
14	Update MFG/PRO customer data with lat	3 days
15	Download cust, add, contact, type ... to e	1 day
16	Contact data (non-MFG/PRO sources)	2 days
17	Update data to Relavis	1 day
18	Upload existing customer activities into Relav	5 days
19	Review business rules	1 day
20	Train users	4 days
21	Configure user profiles and relationships	0.5 days
22	Install Training Database in all users	0.5 days
23	Miller Heiman eLearning	0 days
24	Notes training	1 day
25	Calendaring, sharing calendars, to-do lis	1 day
26	Relavis training	2 days
27	Relavis Product training	1.5 days
28	Business Intelligence	0.1 days
29	Horizon screen-show	0.2 days

With the implementation, sales believe that the whole process needs to be changed. Business Process Maps with the process, key performance indicators (KPIs), responsibilities and systems were drawn up for possible scenarios. After training, in local languages (Japanese, Korean, Mandarin-Chinese), the users were comfortable. A pre-cursor course of general Lotus Notes training was offered to make sure that users were comfortable with functions such as calendaring, to-do lists, etc. An audit of the implementation is planned for the end of the year to find out key success factors and lessons learnt from the implementation. Besides facilitation of documentation about effectiveness of the new system, this audit also provides a baseline measure for future reference. It is best if the audit can provide information about monetary and non-monetary benefits. For example, a balanced scorecard (BSC) approach, a framework developed by Kaplan and Norton, can be adopted. The BSC is organized around four different perspectives: financial; customer (user, or internal customers); internal business processes; and innovation, learning and growth. This approach provides a balance between quantitative and qualitative outcome measures. This project provides company a chance to look for the potentials of virtual office, business process reengineering and knowledge management activities. Knowledge is best to capture in work groups and projects by direct definition by humans, extraction from successful practice, verification and experience [41]. The potential benefits derived here should not be underestimated.

4 Discussion

Whether outcomes are positive or negative, they are likely to change the organizational context in some way. For example, a successful CRM implementation should increase knowledge management capabilities, willingness to share data capabilities and to share data etc. Similarly, an unsuccessful implementation may lead to an opposite effect making staff more reluctant to collaborate or to use the new technology [4]. Sauer's model [42] classifies the list of CRM CSFs (Critical Success Factors) as follows:

- Context: knowledge management capabilities, willingness to share data, willingness to change processes, technological readiness.
- Supports: top management support.
- Project organization: communication of CRM strategy, culture change capability, and systems integration capability.

These three serve to connect the CRM CSFs to the extant body of knowledge on information systems success/failure and to provide a higher-level of abstraction to the CSF list. They also suggest a set of high-level relationships between the CSFs.

Alt and Puschmann [10] applied the following benchmarking procedure (Table 2) to investigate the use of CRM in organizations to identify successful practices. This approach has proved suitable for obtaining information on current practices and results [43]. Alt and Puschmann [10] found that benchmarking showed that CRM involves significant changes regarding the organization of marketing, sales and service activities. Most organizations reorganized internal processes and implemented them on a cross-functional and cross-organizational basis. It is also interesting to know from this study that implementing a CRM system is not mainly driven by the possible savings, 55% of the benchmarked companies agreed that strategic or qualitative goals have been the main drivers for introducing CRM.

Table 2
Summary of benchmarks, criteria and success factors [10]

Benchmarks	Criteria	Critical success factors
Introduction Project	High level of implementation Running CRM system (>6 months)	Start with operational CRM and enhance with analytical and collaborative CRM Rapid evaluation of CRM information systems Medium-term projects which need to be broken down in manageable sub-projects
Organization and customer process	Customer process thinking Analytical CRM (Customer segmentation) Customer centered organization structures	Redesign of customer interaction points and orientation on customer process activities Centralized organization unit for standardization Involvement of top management
System architecture	Centralized customer database Integration of CRM applications Integration of Internet portals	Select CRM system depending on CRM focus Use standard CRM software with minimal customization Integrate systems for analytical and collaborative CRM with operational CRM systems
Efficiency	Quantification of CRM effects Availability of measurement system	Management of projects 'in time' and 'in budget' Measurement of small quantifiable benefits
Culture	CRM as corporate philosophy Availability of change management	Involve users in early stage and communicate CRM goals CRM should not conflict with established organization culture Ensure use of CRM on management level

As mentioned in section 3.4.5 information system audit and balanced scorecard (BSC) approach is underway for comprehensive evaluation. After CRM implementation, team tried to collect information via interviews with key stakeholders and found encouraging results. Quicker turnaround time, reduced internal costs and marketing costs, higher employee productivity and customer retention are some of the benefits as mentioned by stakeholders which will eventually lead to increased revenues and profitability. In terms of intangible benefits stakeholders observed increased customer satisfaction, depth and effectiveness of customer satisfaction, streamlined business processes, closer contact management, improved customer service and better understanding of customer requirements. Therefore this CRM implementation seems successful to a good extent qualitatively in this regard. A questionnaire is under development to measure the effectiveness empirically (validating the CSFs) and report this to all stakeholders as feedback towards further improvement on specific CSF's attributes. The simulation model will also be used for further research to CRM implementation and benefits. The questionnaire will provide data as initial values for the CSF variables in the simulation model. Using these values and other parameters, the simulation can move onward in time in order to explore different scenarios and the consequences of different decisions. This will provide managers with a new and powerful tool with which to exploit the potential of CRM for organizational success. King and Burgess [4] suggested that there is a need for stronger theoretical models of the entire CRM innovation process which can be used by managers to better understand the underlying causes of success and failure. Kim and Kim [44] recommended having an organizational evaluation mechanism to manage, control, and assess the effectiveness of CRM implementation and operational practices. They further argued that a practical perspective based on real experiences as well as theoretical studies is also important to build a framework for measuring CRM performance. CRM is still at an early stage regarding adoption in practice as well as the understanding of success factors in detailed level [10]. They suggested that further research is needed to derive empirically testable hypotheses as suggested by Romano [45] to embed the success factors in a methodology which guides enterprises in successful CRM implementations.

Conclusions

Organizations face considerable challenges in implementing large-scale integrated systems such as ERP and CRM. Implementation of a CRM system was identified as a critical need to align with the overall business strategy of selling solutions, instead of products. The implementation was driven by the business users, with IT playing a facilitating role, thereby making sure that users derive maximum value from implementation. After successful implementation, the CRM system may get into an impact mode, which may challenge business strategy. Various case studies provide different findings which are unique to CRM implementations because of integrative characteristics of CRM systems. As a future plan we would like to

compare various CRM implementations in different organizations on selected significant attributes such as critical success factors and other benchmarks.

Acknowledgement

We would like to thank executive editor and referees for their valuable comments to improve the quality of this paper. We would also like to thank Elzie Cartwright Communicative English Department of Atilim University for nicely editing the manuscript.

References

- [1] Mendoza, L-E., Marius, A., Perez, M., Griman, A-C. (2007) Critical Success Factors for a Customer Relationship Management Strategy, *Information and Software Technology*, 49(2007), pp. 913 -945
- [2] Grönroos, C. (1994) From Marketing Mix to Relationship Marketing: Towards a Paradigm Shift in Marketing, *Management Decision*, Vol. 32, No. 2, pp. 4-20
- [3] Winer, R. S. (2001) A Framework for Customer Relationship Management, *California Management Review*, 43, 4, pp. 89-104
- [4] King, S-F., Burgess, T-F. (2008) Understanding Success and Failure in Customer Relationship Management, *Industrial Marketing Management*, 37(2008), pp. 421-431
- [5] Fingar, P., Kumar, H., Sharma, T. (2000) *Enterprise E-Commerce: The Software Component Breakthrough for Business-to-Business Commerce*. Tampa: Meghan-Kiffer Press
- [6] Shoniregun, C. A., Omoegun, A., Brown-West, D., Logvynovskiy, O. (2004) Can eCRM and Trust Improve eC Customer Base? *Proceedings of the IEEE International Conference on E-Commerce Technology*, IEEE Computer Society
- [7] Szegehgyi, Á. Langanke, U-H (2007) Investigation of the Possibilities for Interdisciplinary Co-Operation by the Use of Knowledge-based Systems, *Acta Polytechnica Hungarica*, 4(2), pp. 63-76
- [8] Reckham, N. (1999) *Rethinking the Sales Force: Redefining Selling to Create and Capture Customer Value*. New York: McGraw-Hill
- [9] Chatham, B., Orlov, L. M., Howard, E., Worthen, B., Coutts, A. (2000) *The Customer Conversation*. Cambridge: Forrester Research, Inc.
- [10] Alt, R., Puschmann, T. (2004) Successful Practices in Customer Relationship Management, 37th Hawaii International Conference on System Science, pp. 1-9
- [11] Gartner survey (2002) *CRM in 2002: Redesign from the customer perspective*. San Jose (CA): Gartner Group, 2001

-
- [12] Thompson, E. (2001) CRM is in its Infancy in Europe. San Jose (CA): Gartner Group, 2001
- [13] AMR Research (2002) The CRM Application Spending Report, 2002-2004, available online at <http://www.amrresearch.com/Content/view.asp?pmillid=10494&docid=9398>
- [14] Verhoef, P. C., Langerak, F. (2002) Eleven Misconceptions about Customer Relationship Management, *Business Strategy Review*, Vol. 13, No. 4, pp. 70-76.
- [15] Davids, M. (1999) How to Avoid the 10 Biggest Mistakes in CRM, *Journal of Business Strategy*, Nov./Dec., 22-26
- [16] Rigby, D. K., Reichheld, F. F., Schefter, P. (2002) Avoid the Four Perils of CRM, *Harvard Business Review*, 80(2), pp. 101-109
- [17] Lindergreen, A., Palmer, R., Vanhamme, J., Wouters, J. (2006) A Relationship-Management Assessment Tool: Questioning, Identifying, and Prioritizing Critical Aspects of Customer Relationships, *Industrial Marketing Management*, 35(1), pp. 51-71
- [18] Ptak, C. A., Scharagenheim, E. (1999) ERP: Tools, Techniques, and Applications for Integrating the Supply Chain, CRC Press-St. Lucie Press
- [19] Wu, J. (2008) Customer Relationship Management in Practice: A Case Study of Hi-Tech Company from China, *International Conference on Service Systems and Service Management*, June 30-July 2, 2008, IEEE Computer Society, pp. 1-6
- [20] Forsyth, R. (2001) Six Major Impediments to Change and How to Overcome Them in CRM in 2001, *Tech. Rep.*, 2001. Available from <http://www.crmguru.com>
- [21] Jancic, Z., Zabkar, V. (2002) Interpersonal vs. Personal Exchanges in Marketing Relationships, *Journal of Marketing Management*, 18, pp. 657-671
- [22] Sheth, J. N., Sisodia, R. S., Sharma, R. S. (2000) The Antecedents and Consequences of Customer-Centric Marketing, *Journal of the Academy of Marketing Science*, 28(1), pp. 55-66
- [23] Morgan, R. M., Hunt, S. D. (1994) The Commitment-Trust Theory of Relationship Marketing, *Journal of Marketing*, 58, pp. 20-38
- [24] Katz, H. (2002) How to Embrace CRM and Make it Succeed in an Organization, SYSPRO white paper, SYSPRO, Costa Mesa, CA.
- [25] Suresh, H. (2004) What is Customer Relationship Management (CRM)? *Supply Chain Planet?*

-
- [26] Hendricks, K. B., Singhal, V. R., Stratman, J. K. (2007) The Impact of Enterprise Systems on Corporate Performance: A Study of ERP, SCM, and CRM System Implementations, *Journal of Operations Management*, 25(2007), pp. 65-82
- [27] Fayerman, M. (2002) Customer Relationship Management. In Serban, A., M., Luan, J. (eds.), *New Directions for Institutional Research, Knowledge: Building a Competitive Advantage in Higher Education*. Chichester: John Wiley & Sons, pp. 57-67
- [28] Bose, R. (2002) Customer Relationship Management: Key Components for IT Success, *Industrial Management and Data Systems*, 102(2), pp. 89-97
- [29] Sebestyenova, J. (2007) Case-based Reasoning in Agent-based Decision-based Decision Support System, *Acta Polytechnica Hungarica*, 4(1), pp. 127-138
- [30] Hahnke, J. (2001) The Critical Phase of the CRM lifecycle. Without CRM Analytics, Your Customer Won't Even Know You're There, www.hyperion.com
- [31] Yu, J. (2008) Customer Relationship Management in Practice: A Case Study of Hi-Tech from China, IEEE Computer Society
- [32] Bull, C. (2003) Strategic Issues in a Customer Relationship Management (CRM) Implementation, *Business Process Management Journal*, 9(5), pp. 592-602
- [33] Gefen, D., Ridings, C. M. (2002) Implementation Team Responsiveness and User Evaluation of Customer Relationship Management: A Quasi-Experimental Design Study of Social Exchange Theory, *Journal of Management Information Systems*, 19(1): 47-69
- [34] Karimi, J., Somers, T. M., Gupta, Y. P. (2001) Impact of Information Technology Management Practices on Customer Service, *Journal of Management Information Systems*, 17(4), pp. 125-158
- [35] Goodhue, D. L., Wixom, B. H., Watson, H. J. (2002) Realizing Business Benefits through CRM: Hitting the Right Target in the Right Way, *MIS Quarterly Executive*, 1(2): 79-94
- [36] Ocker, R. J., Mudambi, S. (2002) Assessing the Readiness of Firms for CRM: A Literature Review and Research Model, *Proceedings of the 36th Hawaii International Conference on System Sciences (HICCS'03)*, IEEE Computer Society
- [37] Yin, R. K. (2003) *Case Study Research: Design and Methods*, 3rd Edition, Sage Publications, Thousands Oaks, CA.
- [38] Feagin, J., Orum, A., Sjoberg, G. (Eds.) (1991) *A Case for Case Study*, University of North Carolina Press, Chapel Hill, NC.

-
- [39] Van Der Blonk, H. (2003) Writing Case Studies in Information Systems Research, *Journal of Information Technology*, Vol. 18, No. 1, March 2003, pp. 45-52
- [40] Mishra, A. Mishra, D. (2009) CRM System Implementation in Multinational Enterprise, R. Meersman, P. Herrero, T. Dillon (Eds.): *OTM 2009 Workshops*, LNCS 5872, pp. 484-493
- [41] Horváth, L., Rudas, I. J., Vaivoda, S., Preitl, Z. (2007) Virtual Space with Enhanced Communication and Knowledge Capabilities, *Acta Polytechnica Hungarica*, 4(3), pp. 17-31
- [42] Sauer, C., Southon, G., Dampney, C. N. G. (1997) Fit, Failure and the House of Horrors: toward a Configurational Theory of IS Project Failure, *Proceedings of the 15th International Conference on Information Systems*, Atlanta, GA, 15-17, pp. 349-366
- [43] Morris, G. W., LoVerde, M. A. (1993) Consortium Surveys, *American Behavioral Scientist*, 36(4), pp. 531-550
- [44] Kim, H-S., Kim, Y-G. (2009) A CRM Performance Measurement Framework: Its Development Process and Application, *Industrial Marketing Management*, 38(2009), pp. 477-489
- [45] Romano, N. C. (2001) Customer Relationship Management Research: An Assessment of Sub Field Development and Maturity. In Sprague, R. H. (ed.), *Proceedings 34th Hawaii International Conference on System Sciences*, Los Alamitos (CA):IEEE

Helical Two-Revolutional Cyclical Surface

Tatiana Olejníková

Department of Applied Mathematics, Civil engineering Faculty
Technical University of Košice
Vysokoškolská 4, 042 00 Košice, Slovakia
e-mail: tatiana.olejnikova@tuke.sk

Abstract: Paper presents a family of helical two-revolutional cyclical surfaces, which are created by movement of the circle alongside the helical cycloidal curve, where circle is located in the curve normal plane and its centre is on this curve. Helical cycloidal curve can be created by simultaneous revolution of a point about two different axes $3o$, $2o$ and by screwing about axis $1o$ in the space. Form of the helical cycloidal curve and also of the helical two-revolutional cyclical surface is dependent on the relative position of the three axes of revolutions, on multiples of angular velocities and orientations of separate revolutions. Analytic representation, classification of surfaces and some of their geometric properties are derived.

Keywords: revolution, angular velocity, cyclical surface

1 Introduction

Helical two-revolutional cyclical surface can be created by movement of the circle alongside the helical cycloidal curve. Circle is located in the normal plane of the curve and its centre is on this curve.

Helical cycloidal curve can be created by simultaneous revolutions of a point about two different lines, axis 3o , 2o and by screwing about axis 1o . Trajectories of the point P, which revolves about single axes of revolutions are circles 2k , 3k located in the planes perpendicular to the axes of revolution 2o , 3o , trajectory of the point P, which screws about axis 1o is helix 1k . With respect to the relative position of axes of revolutions these circles do not necessarily lie in one plane. Form of the helical cycloidal curve is dependent on the relative position of the axes 1o , 2o , 3o , on the orientations of the single revolutions and on their angular velocities, and also on the position of the revolving point P with respect to the axes of revolutions. In the next section there is described the creation of one type of the helical two-revolutional cyclical surface for particular relative position of the axes of revolutions (Figure 1).

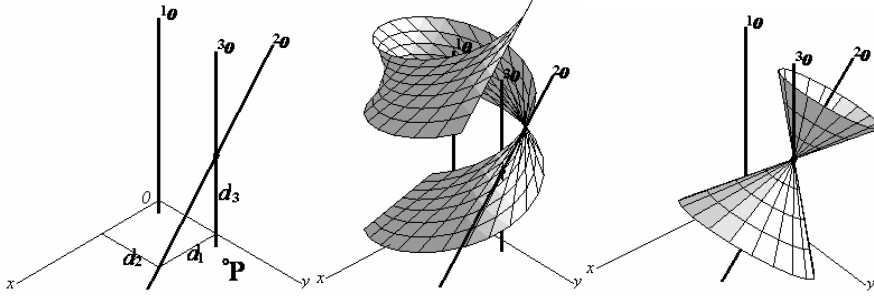


Figure 1
Position of the axes of revolutions

Figure 2
Linear oblique helical surface

Figure 3
Conical surface of revolution

Let axis 1o be fixed and $^1o = z$ in the Cartesian coordinate system (O, x, y, z) . Axis 2o skew to 1o , $^2o \not\parallel ^1o$, creates a linear oblique helical surface by its screwing about axis 1o with angular velocity $w_1 = v$, with orientation determined by parameter q_1 and screw height h (Figure 2). Axis 3o that is intersect to 2o , $^3o \times ^2o$, creates a conical surface of revolution by revolution about axis 2o with angular velocity $w_2 = m_1 w_1 = m_1 v$ and with orientation determined by parameter q_2 (Figure 3). Axis 3o parallel to 1o , $^3o \parallel ^1o$, creates a cylindrical helical surface of revolution by screwing about axis 1o (Figure 4). In Figure 5 there are displayed all three surfaces together. Axis 3o , which revolves about axis 2o and screws about axis 1o simultaneously, creates a composed linear helical-revolutional surface (Figure 6). This surface has four identical branches, because axis 3o revolves about axis 2o with angular velocity, which is 4-multiple of angular velocity of revolution of the axis 2o about axis 1o .

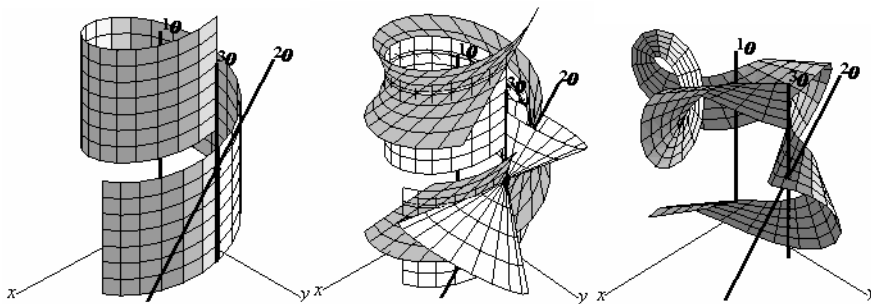


Figure 4
Cylindrical helical surface

Figure 5
All three surfaces together

Figure 6
Composed linear helical-revolutional surface

Point P revolves about axis 3o with angular velocity $w_3 = m_2 w_2 = m_2 m_1 v$ with orientation determined by parameter q_3 , where parameters $q_1, q_2, q_3 = \pm 1$ (if $q_i = +1$, for $i = 1, 2, 3$, then revolution is right-handed, if $q_i = -1$, then revolution is left-handed). Trajectory of the point P movement created by its screwing about axis 1o is helix 1k (Figure 7), the circle 2k is the trajectory of the point P movement about axis 2o (Figure 8) and the circle 3k is the trajectory of the point P movement about axis 3o (Figure 9).

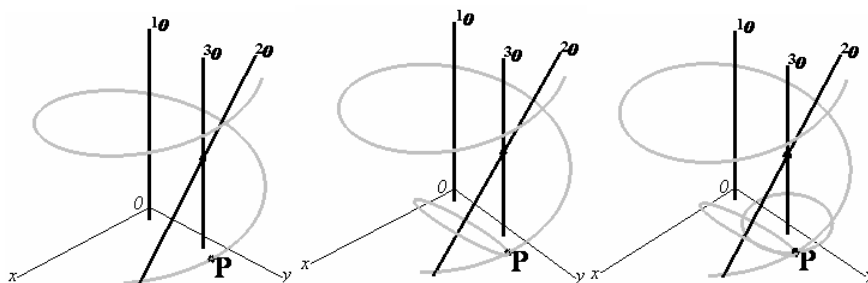


Figure 7
Helix 1k

Figure 8
Helix 1k and circle 2k

Figure 9
Helix 1k and circles $^2k, ^3k$

Curve k as trajectory of the point P composite helical-two-revolutional movement is created by rolling of the circle 3k on the circle 2k , which rolls on the helix 1k simultaneously (Figure 10). Form of this helical cycloidal curve is dependent on the relative position of the axes $^1o, ^2o, ^3o$, on the orientations of the single revolutions and on their angular velocities, and also on the position of the revolving point P with respect to three axes of revolutions.

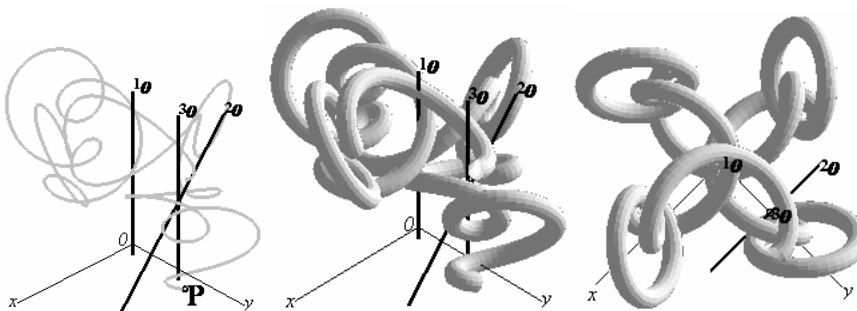


Figure 10
Trajectory of the point P

Figure 11
Helical two-revolutional cyclical surface

Figure 12
View on it from above

Helical two-revolutional cyclical surface can be created by moving a circle alongside the curve k , while the circle lies always in the normal plane of the curve k and its centre is on the curve (Figure 11, in Figure 12 is view from above).

2 Classification of a Family of Helical Two-Revolutional Cyclical Surfaces

Classification of the family of helical two-revolutional cyclical surfaces can be done according to the relative position of axes of revolutions 3o , 2o and 1o , which may be parallel, intersect or skew. Distribution of surfaces within the family is illustrated in the next Figure 13.

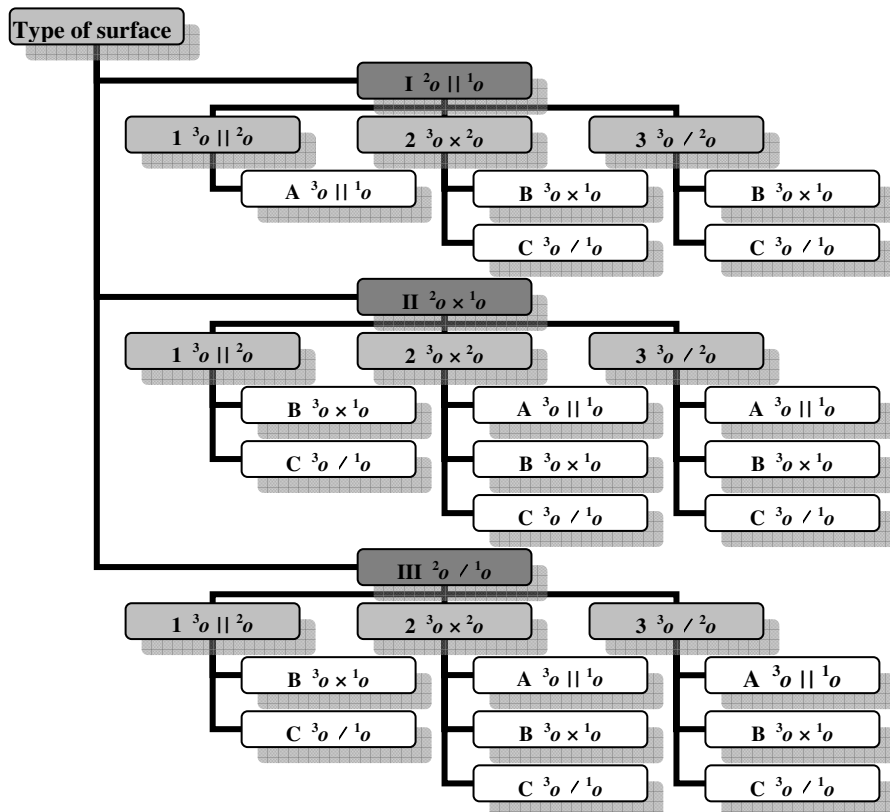


Figure 13

Classification of a family helical two-revolutional surfaces

Helical two-revolutional cyclical surfaces are distributed in the first level into the three types I, II, III with respect to the relative position of the axes 2o and 1o .

Surfaces in all three subclasses I, II, III are distributed in the second level into the three types 1, 2, 3 with respect to the relative position of the axes 3o and 2o .

Finally, in the third level, each subgroup of types 1, 2, 3 can be further classified with respect to the relative position of the axes 3o and 1o into types A, B or C.

3 Analytical Representation of Helical Two-Revolutional Cyclical Surfaces

Let us derive the vector function of the helical two-revolutional cyclical surface for one particular position of the axes of revolutions and for one special position of the point P with respect of these axes, particularly for the surface of type III 2 A. Derivation of the vector function of all other types of surfaces is analogous.

Let the axes of revolution be in the following relative positions: ${}^1o = z$, ${}^2o \parallel {}^1o$ (skew), ${}^3o \times {}^2o$ (intersect), ${}^3o \parallel {}^1o$ (parallel). The position of axis 2o in the plane parallel to the coordinate plane (xz), ${}^2o \subset v'$, $v' \parallel v$, is determined by parameters d_1, d_2, d_3 , which determine the position of the intersection points of axis 2o with the coordinate planes (xy) and (yz) in the Cartesian coordinate system (O, x, y, z). Then $\alpha = \arctg d_3/d_1$ is the angle formed by axis 2o with the coordinate plane (xy) and the position of axis 3o is determined by parameter d_2 , which is the distance between axes 3o and 1o (Figure 1).

Screwing about axis 1o with angular velocity $w_1 = v$, in the direction determined by parameter $q_1 = \pm 1$, with screw height h is represented by matrix

$$\mathbf{T}_1(w_1(v), q_1) = \mathbf{T}_z(w_1, q_1) \cdot \mathbf{T}(0, 0, hv/2\pi), \quad (1)$$

where the matrix $\mathbf{T}_z(w_1, q_1)$ represents revolution about axis z by angle w_1 in the direction determined by parameter q_1 and for $i=1$ it can be derived from (2)

$$\mathbf{T}_z(w_i, q_i) = \begin{pmatrix} \cos w_i & q_i \sin w_i & 0 & 0 \\ -q_i \sin w_i & \cos w_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

and matrix $\mathbf{T}(0, 0, hv/2\pi)$ is translation with vector $(0, 0, hv/2\pi)$ expressed in (6).

Revolution about axis 2o with angular velocity $w_2 = m_1 w_1$, in the direction determined by parameter $q_2 = \pm 1$, is represented by matrix

$$\mathbf{T}_2(w_2(v), q_2) = \mathbf{T}(-d_1, -d_2, 0) \cdot \mathbf{T}_y(\alpha, +1) \cdot \mathbf{T}_x(w_2, q_2) \cdot \mathbf{T}_y(\alpha, -1) \cdot \mathbf{T}(d_1, d_2, 0), \quad (3)$$

where the matrix $\mathbf{T}_y(\alpha, \pm 1)$ expressed in (4) represents the revolution about axis y by angle α in positive or negative direction, matrix $\mathbf{T}_x(w_2, q_2)$ represents revolution about axis x by angle $w_2 = m_1 v$ in the direction determined by parameter q_2 in (5), matrix $\mathbf{T}(\pm d_1, \pm d_2, 0)$ represents translation with translation vector $(\pm d_1, \pm d_2, 0)$ in (6).

$$\mathbf{T}_y(\alpha, \pm 1) = \begin{pmatrix} \cos \alpha & 0 & \pm \sin \alpha & 0 \\ 0 & 1 & 0 & 0 \\ \mp \sin \alpha & 0 & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (4)$$

$$\mathbf{T}_x(w_2, q_2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos w_2 & q_2 \sin w_2 & 0 \\ 0 & -q_2 \sin w_2 & \cos w_2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (5)$$

$$\mathbf{T}(\pm d_i, \pm d_j, \pm d_k) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \pm d_i & \pm d_j & \pm d_k & 1 \end{pmatrix}. \quad (6)$$

Revolutionary movement of the point $P = (x_0, y_0, z_0, 1)$ about axis 3o with angular velocity $w_3 = m_2 w_2 = m_2 m_1 v$ and in the direction determined by parameter $q_3 = \pm 1$ is represented by matrix

$$\mathbf{T}_3(w_3(v), q_3) = \mathbf{T}(0, -d_2, 0) \cdot \mathbf{T}_z(w_3, q_3) \cdot \mathbf{T}(0, d_2, 0), \quad (7)$$

where matrix $\mathbf{T}(0, \pm d_2, 0)$ in (6) represents translation with translation vector $(0, \pm d_2, 0)$, and matrix $\mathbf{T}_z(w_3, q_3)$ is for $i=3$ expressed by (2).

Vector function of the helical cycloidal curve k created by simultaneous revolution of the point $\mathbf{P} = (x_0, y_0, z_0, 1)$ about axes 3o , 2o and screwing about 1o is

$$\mathbf{r}(v) = \mathbf{R} \cdot \mathbf{T}_3(w_3(v), q_3) \cdot \mathbf{T}_2(w_2(v), q_2) \cdot \mathbf{T}_1(w_1(v), q_1), \quad v \in \langle 0, 2\pi \rangle, \quad (8)$$

where $\mathbf{T}_3(w_3(v), q_3)$, $\mathbf{T}_2(w_2(v), q_2)$, $\mathbf{T}_1(w_1(v), q_1)$ are matrices of particular revolutions and screwing expressed in (6), (3), (1) and $\mathbf{R} = (x_0, y_0, z_0, 1)$ is the positioning vector of the point P.

Let the new coordinate system be defined at the arbitrary regular point $P \in k$, identical to the trihedron (P, t, n, b) determined by tangent t , basic normal n and binormal b to the curve k with unit vectors expressed in (9)

$$\mathbf{t}(v) = (t_1, t_2, t_3) = \frac{\mathbf{r}'(v)}{|\mathbf{r}'(v)|}, \quad \mathbf{n}(v) = (n_1, n_2, n_3) = \frac{\mathbf{r}''(v)}{|\mathbf{r}''(v)|}, \quad \mathbf{b}(v) = (b_1, b_2, b_3) = \mathbf{t}(v) \times \mathbf{n}(v). \quad (9)$$

Helical two-revolutional cyclical surface can be created by movement of the circle $c = (P, r)$ with centre P and radius r alongside the curve k so that the circle is located in the normal plane of the curve in the point $P \in k$, which is determined by basic normal n and binormal b to this curve. Vector function of this surface is

$$\mathbf{P}(u, v) = \mathbf{r}(v) + (n_1 r \cos u + b_1 r \sin u, n_2 r \cos u + b_2 r \sin u, n_3 r \cos u + b_3 r \sin u), \quad (10)$$

for $u \in \langle 0, 2\pi \rangle$, $v \in \langle 0, 2\pi \rangle$, where $\mathbf{r}(v)$ is vector function of the helical cycloidal curve k expressed in (8).

Form of the helical cycloidal curve k and created helical two-revolutional cyclical surface changes in dependence on the relative position of the axes of revolutions that are determined by parameters d_i , $i = 1, 2, 3$. Surface has m_1 identical external branches, where every branch has m_2 identical internal branches. Point P revolves about axis 3o with angular velocity w_3 , which is m_2 -multiple of the angular velocity w_2 of the revolution about axis 2o and w_2 is m_1 -multiple of the angular velocity w_1 of the revolution about axis 1o . Many different forms of cycloidal cyclical surfaces can be created by change of their determining parameters.

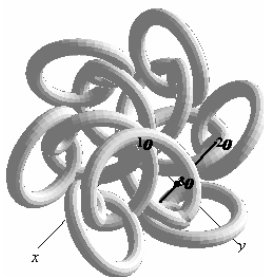


Figure 14
 $m_1 = 6$, $m_2 = 2$

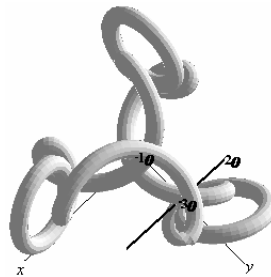


Figure 15
 $m_1 = 3$, $m_2 = 2$

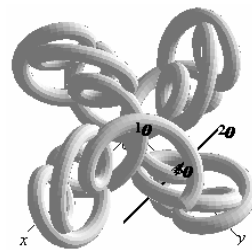


Figure 16
 $m_1 = 4$, $m_2 = 3$

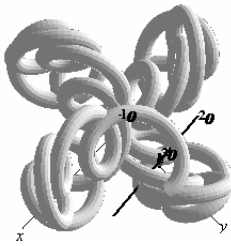


Figure 17

$$m_1 = 4, m_2 = 4$$

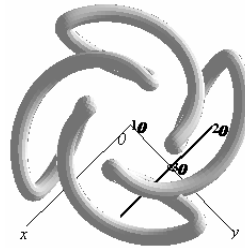


Figure 18

$$m_1 = 4, m_2 = 2, q_2 = -1, q_3 = -1$$

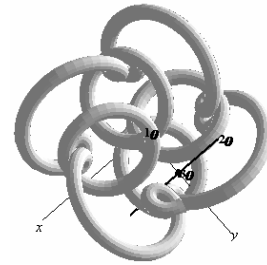


Figure 19

$$m_1 = 4, m_2 = 2, q_2 = -1, q_3 = +1$$

Variations of the surface form are shown by change of some parameters of the surface of type III 2 A displayed in Figures 11 and 12. Presented surface is determined by parameters $m_1 = 4, m_2 = 2, q_1 = q_2 = +1, q_3 = -1$, then it has 4 external and 2 internal branches, and all three revolutions are not right-handed. Surface in Figure 14 is determined by parameter $m_1 = 6, m_2 = 2$, in Figure 15 by $m_1 = 3, m_2 = 2$, in Figure 16 by $m_1 = 4, m_2 = 3$, in Figure 17 by $m_1 = 4, m_2 = 4$, then there are changes in the number of external and internal branches.

In Figure 18 depicted surface is determined by parameters $m_1 = 4, m_2 = 2, q_2 = -1, q_3 = -1$, in Figure 19 by $q_2 = -1$ and $q_3 = +1$, then there are changes in the orientations of particular revolutions.

In Figure 20, there is presented surface with parameters identical to parameters of surface in Figures 11 and 12, but the position of the point $P(x_0, y_0, z_0, 1)$ was changed from $(d_1/2, d_2/2, 0, 1)$ to $(d_1, 0, 0, 1)$.

Surface with parameters $m_1 = 4, m_2 = 6, q_2 = -1, q_3 = +1$ is illustrated in Figures 21 and 22, but relative position of the axes has been changed to position ${}^2o \perp {}^1o$, ${}^2o \perp {}^3o$ and ${}^2o \perp {}^3o$. Surfaces in Figures 14-21 are displayed by view from above, because in these views the changes of parameters are more illustrative.

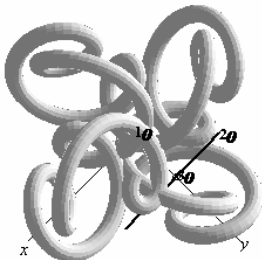


Figure 20

$$m_1 = 4, m_2 = 6, q_2 = -1, q_3 = +1$$

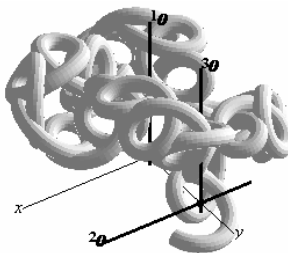


Figure 21

New position of the point **P**

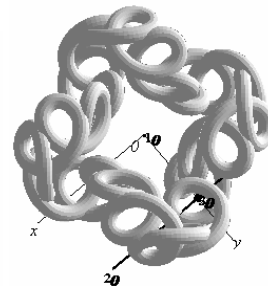
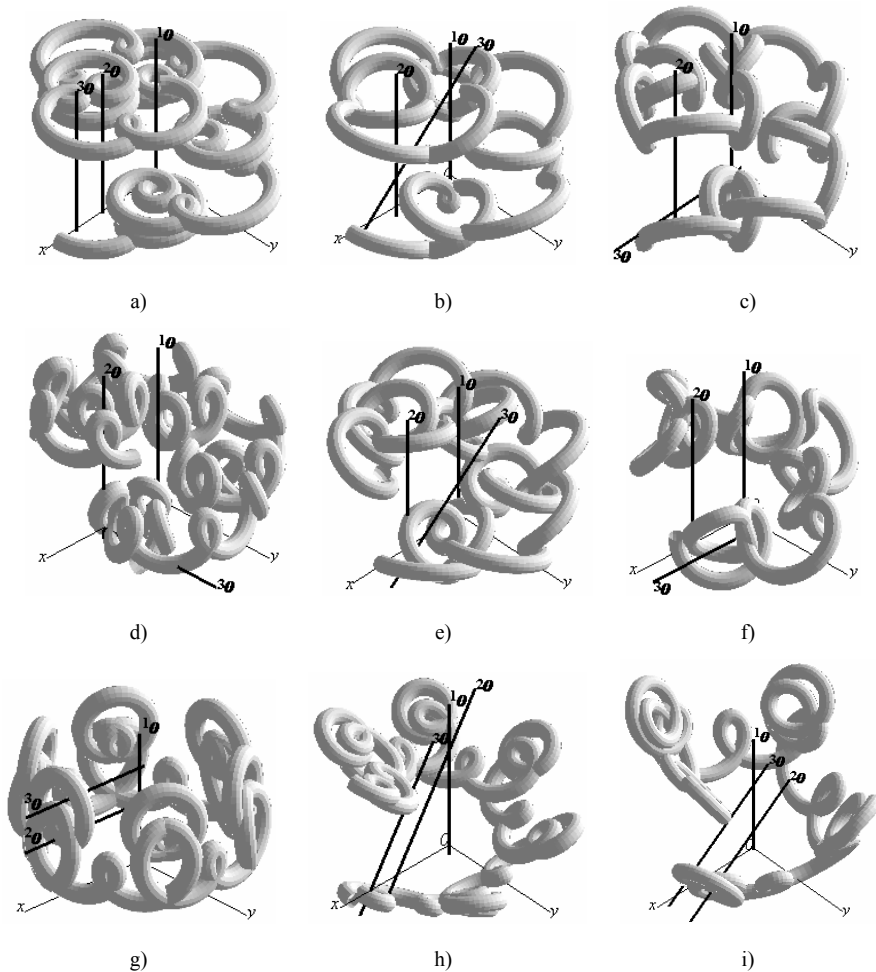


Figure 22

View on it from above

Conclusion

As the conclusion it can be summarised that the presented family of helical two-revolutional cyclical surfaces serves as an endlessly rich source of inspiration for artistic and design purposes. Their unusually complex forms obtained in a relatively simple way of composite spatial transformation. Special skew symmetry and harmonical periodicity reflect their simplistic generating principle based on the naturally basic movement of our universe, revolution about an axis in the space. Several surface types from the presented classification frame are displayed in the Figures 23 a)-o) without commentary, as the most persuasive evidence.



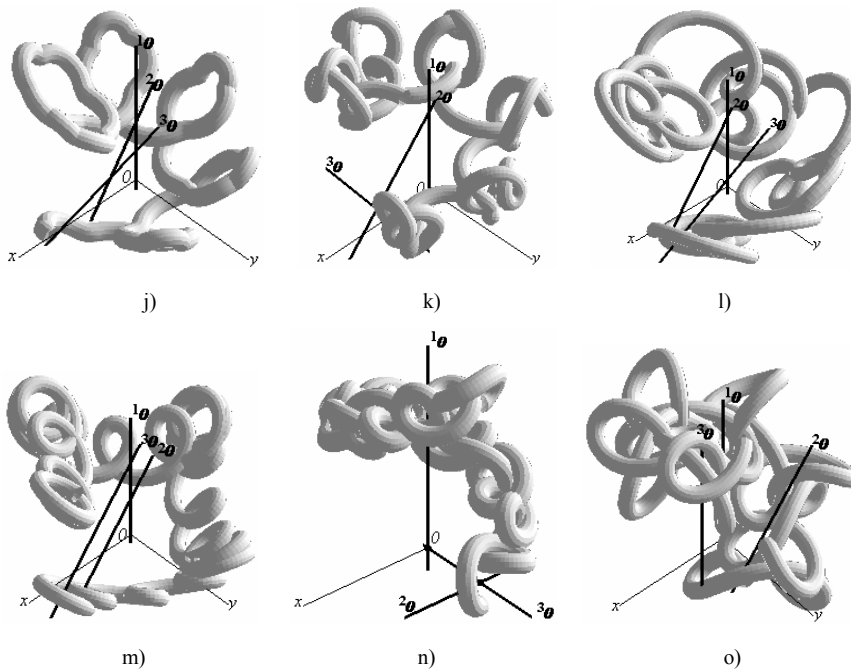


Figure 23

Surface types from the presented classification frame

Acknowledgement

This work was supported by the VEGA 1 / 4002 / 07 “Surfaces in geometrical modelling” and by project OPVaV-2008/2.1/01 “Support of Centre of integrated Research of progressive building constructions, materials and Technologies”.

References

- [1] B. Budinský, B. Kepr: Basic of Differential Geometry with Technical Applications, SNTL-Publishers of technical Literature, Praha, 1970
- [2] L. Granát, H. Sechovský: Computer Graphics, SNTL-Publishers of technical Literature, Praha, 1980
- [3] T. Olejníková: 3D Cycloidal Curves and Cyclical Surfaces, in: Proceedings of International Conference „70 years of SvF STU“, Section 05, Bratislava, Slovakia, 2008
- [4] E. Stanová: Composition of Helical and Cycloidal Motions, Proceedings of “14. conference for descriptive Geometry, computer Graphics and technical Draving”, Bílá, Czech Republic, 1994, pp. 67-72
- [5] D. Velichová: Trajectories of Composite Revolutionary Movements, in: G, Slovak Journal for Geometry and Graphics, Sjf STU Bratislava, Slovakia, 2006, pp. 47-64

Behaviour Study of a Multi-Agent Mobile Robot System during Potential Field Building

István Nagy

Institute of Mechatronics and Vehicle Techniques
Bánki Donát Faculty of Mechanical and Safety Engineering
Budapest Tech
Népszínház u. 8, 1081 Budapest, Hungary
nagy.istvan@bkgk.bmf.hu

Abstract: In this paper a multi-agent based mobile robot simulation system will be presented where the behaviour of the system is studied with different number of agents (1, 3,6) and also with different number of ultrasonic range sensors on agents (8 or 16 US sensors on individual agents). The task of the autonomous agents is to create the potential field (PF) of an unknown environment. The classic problems of PF building, like oscillation and trapping, are not the focus of the article, but instead, the article is concerned with the agents' self-organizing ability where self-organizing is controlled by a genetic algorithm (GA). The GA is equipped with two fitness functions where one "maintains" the distances between certain agents (spat distr), while another "watches" the area coverage (area cover). In fact, the paper can be divided into three main parts. The first part describes the ultrasonic sensing and range measuring with systematic errors, the potential field (PF) building and the moving strategies. The second part contains description of the GA, the operation of the GA, the structure of the system, the fitness functions and a general system-error determination. In the final third part, the obtained results are analyzed and presented in the appendices.

Keywords: Genetic algorithm (GA), Mobile Robot, Multi-agent, Potential Field

1 Aims and Motivation

Nowadays, in mobile robot research a huge amount of literature is available about path planning and course controlling based on a potential field. These articles are mostly about eliminating or preventing the classic problems arising in potential field building, such as trapping and oscillation. With the evolution of this area of knowledge, newer and newer methods are appearing for handling these mentioned problems, but these methods usually concern single agents. A good example can

be seen in [1], where the authors are eliminating the oscillation problem by a VFB¹ guiding model, in which at path planning the VFB is realized by a neuro-fuzzy model producing an oscillation free path between the starting and docking positions. The developed algorithm was tested in a virtual training environment named “COSMOS”, and the results can be found in the mentioned article. In relation to this, another example can be mentioned where the classic parking problem is realized by a hybrid navigation structure, with the elements of computational intelligence [2]. The hybrid structure has three components: *harmonic PF* (calculation of the path in an initial – static – environment); *neural network* (trying to control the robot to pass through the orientation marks that the path is composed of); *fuzzy controller* (obstacle avoidance and trying to find the next orientation marks again). For simulation results see [2].

My primary aim is to create a functional simulation system that will be able to create the potential field of an unknown environment on a multi-agent platform. While developing it I will not devote to the classic problems of potential field building (trapping, oscillation), but I rather wish to control the group behaviour of agents, believing that the previously mentioned basic problems can also be eliminated by this. My secondary aim is, in case of a successful system, to accomplish its analysis (see conclusion) and (probably in another article) to increase the efficiency of the algorithm by tuning the system or the GA parameters. Later in the future I would like to apply this algorithm for multi-agent systems with different sensors (e.g. visual sensors) as well.

2 Introduction

This paper actually is a continuation of the conference paper [3], and this is why the basic definitions and determinations published previously are mentioned here only in a shortened form.

Distributed problem solving at multi-agent mobile robot systems has its origin in the late 1980s [4], [5], however, since 2000, the field of cooperative mobile agents has shown dramatic development. It is reasonable to ask: Why should we use multi-agent mobile robot systems? Answering it, let me compare several advantages of multi-agent systems, as contrasted with single-agent ones.

- More efficiency (faster and more accurate).

Keeping to the main topic of the paper, in multi-agent systems – by exchanging the main information between one another –, the individual agents are capable to localize themselves faster and more accurately.

¹ *Vector Field Based guiding model*

- More fault-tolerant.

Namely, if in a single-agent system the agent breaks down, the task will not be executed, while in a multi-agent system, though depending on its intelligence, the execution of the task is continued.

Generally speaking, a multi-robot system has the following remarkable properties:

- Larger range of task domains (flexibility)
- Fault-tolerance
- Greater efficiency, robustness

In the development of multi robot systems, primary merits can be attributed to M. J. Mataric (MIT, USA) whose scientific achievements include researching and developing strategies of behaviour-based mobile robot systems [6], [7], [8]. Each of these studies contains relevant statements and definitions in the field of individual or group behaviour of mobile robots. The individual agent is very well defined by Tecuci in [9] –“the agent is an autonomously active entity with certain possibilities to sense its environment and act in it in order to achieve certain states of this environment in which certain previously specified goals are achieved”. Later, by the development of this field of science different types of agents were defined, and this can be observed very well in [10], where the basic classification of agents is extended and apart from this the agents are classified from a functional-computational perspective. After the definition of single agents, we can now focus on multi-agent systems and mainly on the cooperation between individual agents. In [11], the authors try to draw the agents into an agent coalition for the sake of a more efficient task execution. Firstly, the agent coalition is formed, the individual agents are rated and some value is assigned to them. Then, based on the agent’s value, the agent will join the coalition if the coalition brings to the agent at least the same or better results than when it works independently. Another important contribution has been made by Fukuda and Iritani, who tried to widen the possibilities of the cooperation between separated agents in multi agent mobile robot systems [12].

The simulation system, described in this paper has a modular structure. There is a separate module for the sensory system of the agents (which is the mathematical model of the ultrasonic range detector), another module contains the GA, responsible for near-optimal behaviour selection, and the next separated module is responsible for displaying results and assessments.

Since the simulation system has been prepared in a MATLAB environment, it is inevitable to make the mathematical model of the system. The workspace is digitally decomposed (grid construction), and the agents are point-represented in this model. The visited areas are renumbered during the process of map-building, in order to avoid duplicity of the map occurring in the same area. The potential field building and calculation is based on the principle of the well-known

repulsive forces. A simplified map building process by one agent is represented on Figure 1. The agent moves to the new position, assigns the position to the grid construction of the model, performs distance measurements, evaluates the potential field value (broadcasting the parameters of its own new position to the other agents), and then plans the next move. The potential field values are stored on the host remote server, where the global potential field map of the whole WS will be updated. In the advanced systems (will be represented in this paper), in order to avoid collisions, the moving mechanism is controlled by the GA.

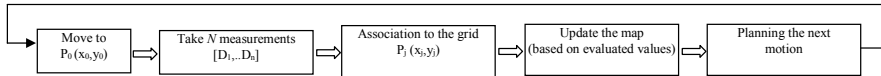


Figure 1

A simple map building process

3 Sensing

The individual agents are equipped with 8 or 16 ultrasonic sensors for distance measurements. The sensors are equally spaced on a ring around the body of the robot (see Figure 2a) so that the sensors form a regular octagon (or a polygon with 16 points) on the circle of the agent. In this case the sensing sector of each sensor can be calculated with the form:

$$\beta = \frac{2\pi}{N}; \quad (1)$$

where, N is the number of sensors. The sensors can also choose either *long-* or *short-range* sensing. The long-range sensing (*LRS*) perceives the obstacle or other agents in the given sector (β) in infinite² distances. The short-range sensing (*SRS*) is determined in a circle with radius R_0 . Occupation of the segments by other agents or obstacles, is represented with a binary word, and will have importance in choosing the next behavior or the moving mechanism.

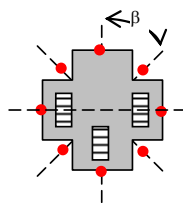


Figure 2a

The agent, and the sensors around, located by angle β

² infinite=beyond the given radius R_0

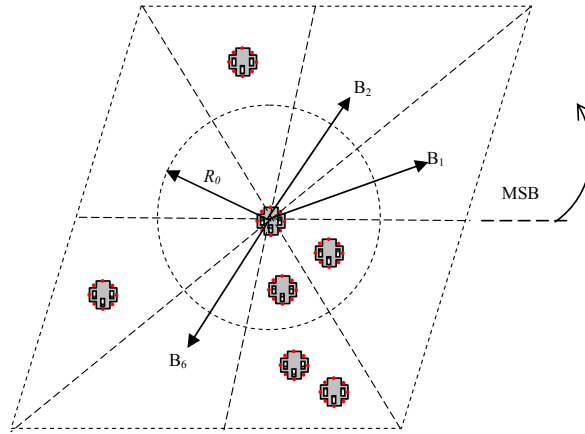


Figure 2b

The Long, and Short range sensing

Let us see the illustration given on Figure 2b, where the binary words of short-respectively long-range sensing are:

LRS: 00101010

SRS: 00000011

Mathematically can be written:

$$b_j = \begin{cases} 1, & \exists P_i \in \varepsilon_j; \\ 0, & \forall P_i \notin \varepsilon_j \end{cases}; \quad (2)$$

where, b_j is the value, given by the j^{th} sensor, and P_i is the position of i^{th} agent in sector ε_j . In case of short-range sensing the sensing sector (ε_j) is valid only in the given R_0 radius.

$$\underline{\varepsilon}_j = \{P_i \mid |P_i - P_0| < R_0\}; \quad (3)$$

where, P_i is the position of i^{th} agent, and P_0 is the position of reference agent [13]. The surrounding environment of the reference robot is represented with the binary words *LRS* and *SRS*. We can say that two binary words are equivalent if the number of 1s and the position of 1s in relation to one another are identical (e. g. the words $\zeta_1=01100000$ és $\zeta_2=00000110$ are equivalent). It is observable that with shifting to left or right, or with circular operations we can get several equivalent words. Let us name these equivalent bits *stimulus* and label (ζ). The stimulus contains the description of the environment of the mobile robot [13].

In the perception model, the starting positions of the agents are already known (see Appendix 3). In an ideal case, the (d) distance is calculated from the time of

flight (t) and the spreading of speed of sound (v), in case of ultrasonic range measurements [14].

$$d = \frac{1}{2} v.t; \quad (4)$$

In this model the ideal case is considered, that is after checking the sensing segment's occupation, the distance calculations in x and y directions has been provided. The distance measurement model can be seen on Figure 3.

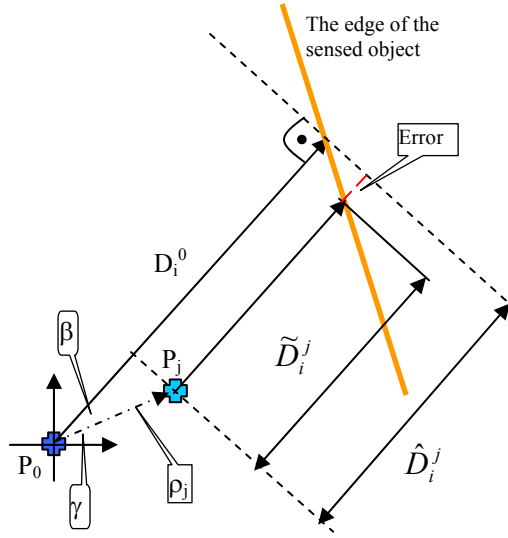


Figure 3

The mathematical model of *US* sensing

As a result of *WS* rasterizing (grid construction), an exact result of distance measurement is almost impossible. Unfortunately this is inconceivable in real environment, since we have to take into consideration the error (δ_D), see Figure 3.

$$\delta_D = |\tilde{D}_i^j - \hat{D}_i^j|; \quad (5)$$

where, \tilde{D}_i^j is the real distance, and \hat{D}_i^j is the evaluated one.

$$\hat{D}_i^j = D_i^0 - \rho_j \cdot \cos \beta, \quad i = 1..8, \text{ resp. } (1..16); \quad (6)$$

The distances measured with N ultrasonic sensors are stored in the L measuring-vector. The measuring-vector belonging to the P_0 location is: $L_0 \equiv [D_1^0, D_2^0, \dots, D_8^0]$. Besides, the evaluated distances in the model (after the grid association) belonging to the P_j position are stored in the distance-vector: $\hat{L}_j = [\hat{D}_1^j, \hat{D}_2^j, \dots, \hat{D}_8^j]$. The errors depend on the complexity of the environment and certainly on the map grid

width (see Figure 3). With reference to the members of L_j vector, for the sake of better evaluation results, the weighting vector (w_i) has been introduced. Regarding the distance between the robot location and P_j raster position, the weighting can be written as:

$$w_j = e^{-\eta \rho_j^2}; \quad (7)$$

where η is a positive constant, and ρ_j is the distance between P_0 - P_j locations (See Figure 3). Namely $w_j=1$ if the agent is exactly in the position P_j (in this case $P_j=P_0$).

4 The Potential Field

The creation of the artificial potential field (APF) has been done by the well-known repulsive force method [1], [13].

$$\begin{aligned} U_{ART}(x) &= U_{GOAL}(x) + U_{OBS}(x) \\ U_{GOAL}(x) &= -\frac{1}{2}k_p(x - x_{GOAL})^2 \\ U_{OBS}(x) &= \begin{cases} \frac{1}{2}\eta \left(\frac{1}{x} - \frac{1}{l_0}\right)^2; & \text{if } x \leq l_0; \\ 0 & \text{if } x > l_0; \end{cases} \end{aligned} \quad (8)$$

$$\vec{F}_{ART} = -\nabla[U_{ART}(x)];$$

where, U_{ART} is the APF, U_{GOAL} is the potential field spreading from the goal, U_{OBS} is the potential field of the obstacle, k_p is a positive gain, l_0 is a threshold limit beyond which are no repulsive forces, and η is a positive constant.

The potential field building

In case of validity of the next condition:

$$\forall i \in [1, N], \quad \hat{D}_i^j \geq 0; \quad (9)$$

the potential field is calculated from the \hat{L}_j vector. This condition is valid for the visibility of P_j position simultaneously. If the above mentioned condition is not valid, it means that the agent is on the obstacle, or is part of the obstacle. The evaluated value of the potential field at the location P_j in step “ t ” (if the above mentioned condition is valid), is:

$$\hat{U}_j^t \cong \sum_{i=1}^N e^{-\lambda \hat{D}_i^t}; \quad (10)$$

where λ is a positive coefficient. The potential field values, belonging to P_j position, measured by k^{th} mobile robot, at time “ t ”, are stored in set Ω .

$$\Omega_j^t = \{\hat{U}_j^{t1}, \hat{U}_j^{t2}, \dots, \hat{U}_j^{tk}\}; \quad (11)$$

In case if a member of Ω_j^t set equals zero, then is valid:

$$\begin{aligned} \Omega_j^t &= \Omega_j^{t-1} \cup \Theta; \\ \Theta &= \begin{cases} \hat{U}_j^{tk}, & \text{if for } \hat{L}_j - \text{is valid (9);} \\ 0, & \text{otherwise;} \end{cases} \end{aligned} \quad (12)$$

To the Ω_j^t set, is associated the following confidence weight vector:

$$W_j^t = \{w_j^{t1}, w_j^{t2}, \dots, w_j^{tk}\}; \quad (13)$$

where the normalized weight component of W_j^t is:

$$w_j^{-ti} = \frac{w_j^{ti}}{\sum_{n=1}^k w_j^{tn}}; \quad (14)$$

Finally an acceptable potential field value can be readily calculated as follows:

$$U_j^t = \begin{cases} \hat{U}_j^u, & \text{if } \exists i \in [1, k], w_j^u = 1; \\ \text{else: } \sum_{i=1}^k \hat{U}_j^u \cdot w_j^{-ti}; \end{cases} \quad (15)$$

5 Motion Mechanism

After the execution of sensing, measurements, and estimating the value of the potential field in position P_0 , the agent has to move to the next position to continue its measurements. This move can be applied as based on three motion selection [13]:

Directional1 – here the standard deviation of potential field is calculated in all (N) sensing sectors within the given maximum movement step (dm) at time “ t ” and “ $t-1$ ”. Moreover, the move in time “ $t+1$ ” will be calculated according to the motion direction (ϕ) and motion step (d_s). For the P_0 location of the robot at time “ $t+1$ ” can be written:

$$P_0^{t+1} = P_0^t + d_s \cdot e^{j\phi}; \quad (16)$$

Let us store the difference of the standard deviations of potential fields at the same sensing vector, at time “ t ” and “ $t-1$ ” in vector Δ . Then the i^{th} component of this vector is:

$$\Delta_i = \text{std}(\{l_{ij} \mid l_{ij} = U_{ij}^t - U_{ij}^{t-1}, \forall j \in \varepsilon_i, i = 1, 2, \dots, N\}); \quad (17)$$

Besides, let array Λ be the standard deviation of potential field for all locations in the same sensing sector at time “ t ”. For the i^{th} component of this vector can be written:

$$\Lambda_i = \text{std}(\{v_{ij} \mid v_{ij} = U_{ij}^t, \forall j \in \varepsilon_v, |j - i| \leq 1\}); \quad (18)$$

where ε_v is defined similarly like ε_i - see above. After that, the motion direction in sector i :

$$\begin{aligned} \phi_i \mid \Delta_i &= \max(\Delta_1, \Delta_2, \dots, \Delta_N); \\ \forall j, P_j &\notin \varepsilon_i; \end{aligned} \quad (19)$$

where P_j is the location, and ε_i is the sensing sector. Namely, the agent will select its direction of movement (ϕ) in “ t ” sector, based on the condition (19). The exact position, $P_0^{t+1}(x_0, y_0)$, within the selected sector, should satisfy the following condition:

$$(x_0, y_0) \mid \Lambda_i(x_0, y_0) = \max(\Lambda_1, \Lambda_2, \dots); \quad (20)$$

Directional2 – The strategy is almost the same as previously (see *Directional1*), the only difference being in selecting the exact position within the selected sector. The exact position selection is based on the minimum value of vector Λ .

$$(x_0, y_0) \mid \Lambda_i(x_0, y_0) = \min(\Lambda_1, \Lambda_2, \dots); \quad (21)$$

Limited random – The agent selects its motion direction and step size randomly, within the given limits.

$$\phi_i = \text{rand}([1..N]); \quad (22)$$

$$d_s = \text{rand}([1..d_m]); \quad (23)$$

The next question should be about how to make the potential field building process more effective. The answer lies in the appropriate behaviour mechanism.

6 Behaviour Selection Mechanism

Usually, in behaviour-based task execution at mobile robots, the next behaviour is very much influenced by the environment (as we know, the environment of the mobile robot is represented by stimuli; see above). In a general case, it exists as a set of behaviours out of which the best behaviour will be chosen by the algorithm, based on the environment's appraised measurements.

Behaviour in single agent environment – Let the primitive behaviour correspond to the direction of the 8 (or 16) ultrasonic sensors (see Figure 2b, $\{B_1, B_2, \dots, B_8\}$). In other words, the elementary motions are summarized in vector B , $B=[B_1..B_N]$. The values of this vector can be $B_i \in \{-1, 1\}$ in such a way that: $B_i=1$, if the agent is capable of executing the required move, else $B_i=-1$. The behaviour and the weighting vectors are:

$$B = \begin{bmatrix} B_1 \\ B_2 \\ \cdot \\ \cdot \\ B_N \end{bmatrix}; W = \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ w_N \end{bmatrix}; \quad (24)$$

where, the values of weighting: $W_i=-1$, if $B_i=-1$, and in other cases the weighting is:

$$\sum_{i=1}^N W_i |_{w_i \neq -1} = 1; \quad (25)$$

The simplified behaviour selection process can be summed up as follows: Based on the sensing vector (see *LRS*, *SRS* mentioned above), the appropriate stimulus is given which triggers a condition for the behaviour selection mechanism. Next, as the output of the embedded learning mechanism (See Figure 4; dashed line), the near optimal behaviour will be selected. Let us have a few more words about the simplified learning mechanism. Any response to the sensing vector of the agent (what is nothing else than a stimulus, and the response to the stimulus, which is the behaviour) is represented by the varying weight. The learning step is the following: if the agent has selected the motion direction, then the components of the W vector will be updated. As a result of a series of updating, the outcome will be some more significant directions.

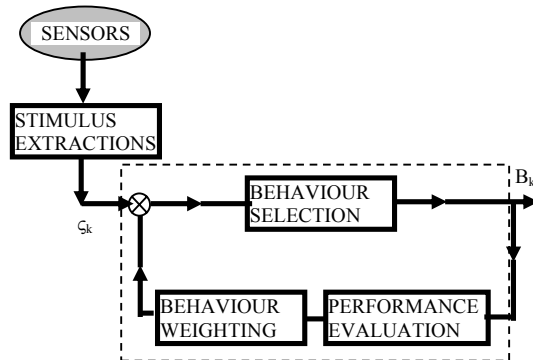


Figure 4

The simplified behaviour selection

The process: a stimulus (ζ_k) is chosen, to which the appropriate behaviour (B_k) belongs, next the motion is executed, and then the operation is appraised by weighting (W). The next stimulus-behaviour pair selection is based on this weighting vector, which produces a more effective motion mechanism.

Behaviour in a multi-agent environment – Let the agent to the stimulus (ζ_k) select the behaviour (B_k), at time „ t ”. After it, the robot learns, based on its local performance criteria. In the case if a common basis for behaviour selection is used, the agents can share their learned knowledge. The behaviour weight vector will be updated, based on the following:

$$W_{\zeta^k}^{t+1} = \text{normal}(\text{shape}(W_{\zeta^k}^t + \Delta W)); \quad (26)$$

where operator “*normal*” is normalizing the weight vector, and ΔW is an increment vector. The performance of operator “*shape*” is illustrated on Figure 5, where the updated weight vector passes through *function1*, and conditionally *function2* [15].

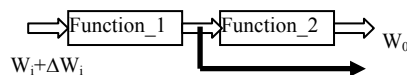


Figure 5

The “shape” operator

The *function1*:

$$w_0 = \begin{cases} 0, & \text{if } w_i < 0 \\ w_i, & \text{if } 0 \leq w_i \leq 1; \\ 1, & \text{if } w_i > 1 \end{cases} \quad (27)$$

The *function2*:

$$w_0 = \frac{\alpha}{1 + e^{-w_i}} - \psi; \quad (28)$$

where coefficients α and ψ influence the shape of the function. Then, the j^{th} component of the ΔW weight-increment vector is:

$$\Delta w_j = \begin{cases} \delta |_{E(B_k)}, & \text{if } j = k \\ 0, & \text{if } j \neq k \end{cases}; \quad (29)$$

where $E(B_k)$ is an evaluation of behaviour B_k and $\delta \in [-1, 1]$. At time $t=0$, the i^{th} component of the behaviour weight vector is:

$$w_i^0 = \begin{cases} -1, & \text{if } B_i = -1 \\ \frac{1}{\beta}, & \text{otherwise} \end{cases}; \quad (30)$$

where β is the number of the feasible behaviours.

Behaviour selection mechanism – this mechanism assigns the behaviour weight to the corresponding behaviour (*sel*: $W \rightarrow B_k$). The behaviour selection mechanism can work in two ways:

a) selection based on the *probability of the behaviour weight vector distribution*:

$$B_{\text{sel}} = B_k |_{P(w_k)}; \quad (31)$$

b) selection based on maximum weight:

$$B_{\text{sel}} = B_k |_{w_k = \max(w_1, w_2, \dots, w_N)}; \quad (32)$$

After behaviour selection the agent moves along in the selected direction, with step size d_0 . Let us mark the position at time “ t ” $\rightarrow P^t$, and at time “ $t+1$ ” $\rightarrow P^{t+1}$. In this case this whole process (action) can be defined as [15]:

$$P_i^{t+1} = \text{action}(B_k, d_0, P_i^t); \quad (33)$$

The next step in behaviour based robotics was, mainly in environments where multi-agent robot groups occur, that for the selecting of near optimal behaviour genetic algorithms and/or neural networks were used. In this paper the near optimal behaviour is selected through a genetic algorithm which is working with two fitness functions. The essence of behaviour-control is that the agents are organized into robot groups, the efficiency of the individual agents is evaluated, and then based on this evaluation the next “action” is selected. In “action”, the direction selection is considered with two situations. The first is the *spatial distribution* of agents (when the distance between two agents i and j is less than

the given threshold distance: $d_{ij} \leq R_2$). The second is the *area coverage*, when $d_{ij} > R_2$.

Spatial distribution - For the reference robot i , and m – neighbouring robots, is valid:

$$i, \forall m \in [1, M], m \neq i, d_{im} \leq R_2;$$

$$\text{spat}_{m=1}^{m_d-1} \text{distr} \frac{e^{-j\gamma_m}}{d_{im}} \cong \xi_i e^{j\theta_i}; \quad (34)$$

Area coverage:

$$i, \forall m \in [1, M], m \neq i, d_{im} > R_2;$$

$$\text{area}_{n=1}^N \text{cov er} \frac{e^{\frac{j2\pi n}{N}}}{D_n^0} \cong \xi_i e^{j\theta_i} \quad (35)$$

where, d_{im} is the distance between robot i and m , then m_d is the number of group robots inside of R_2 threshold limit, γ_m is the relative angle of motion direction (see Figure 3, where m^{th} agent is moving to P_j location and $\theta_i \neq \rho_i$), and D_n^0 is the n^{th} component of L_0 vector.

Let the significant proximity direction a time t be θ_i^u (that is the direction of the i^{th} agent's motion is u , where u is one of the 8/16 sensing sectors: $u \in [1..N]$). There exists a probability vector ϖ_i , where the components express the efficiency of (34) and/or (35) if the motion was executed. This ϖ_i vector can be written as follows:

$$\varpi_i = [\phi_1, \phi_2, \dots, \phi_N];$$

where $\phi_k \in [0, 1]$ and $\sum_{k=1}^N \phi_k = 1$; . In case if the agent in the next motion (at time $t+1$) selects a different motion direction “ v ”, ($v \in [1..N]$), denotes it by θ_i^v , then the k^{th} component of the ϖ_i vector will be updated as follows:

$$\phi_k^{t+1} = \frac{\phi_k^t + \delta}{1 + \psi}; \quad (37)$$

where

$$\delta = \begin{cases} \psi, & \text{if } k = v - u + 1 \\ 0 & \text{otherwise} \end{cases};$$

where ψ is a positive coefficient. As a result of permanent updating some motion directions become more significant than others.

7 Genetic Algorithm

In this system the near optimal motion direction is selected through a GA. The simplified operation of the genetic algorithm works as follows:

- The fitness of each member in a GA population is calculated according to an evaluation function (*fitness functions*), which measures how well the individual performs.
- Individuals performing well are propagated in proportion to their fitness; on the other hand, the poorly performing members are reduced or eliminated completely.
- By exchanging the information between members it is possible to create new search points, by which the population explores the search space and converges in an optimal solution.

To find and represent these new search points, the GA uses its operators. Several operators are known, but the three most frequently used ones are: *reproduction* (selects the fittest members and copies them exactly; *crossover* (swapping some part of their representations.); *mutation* (prevents the loss of information that occurs as the population converges on the fittest individuals).

In every step the mobile robot checks its environment, then according to the vectors (34), (35) and the probability vector ϖ_i , (what is the result of the learning process), next motion direction is selected. In compliance with this probability vector, the GA *population* will be determined on the basis of this ϖ_i vector. The structure of this whole system can be seen on Figure. 6.

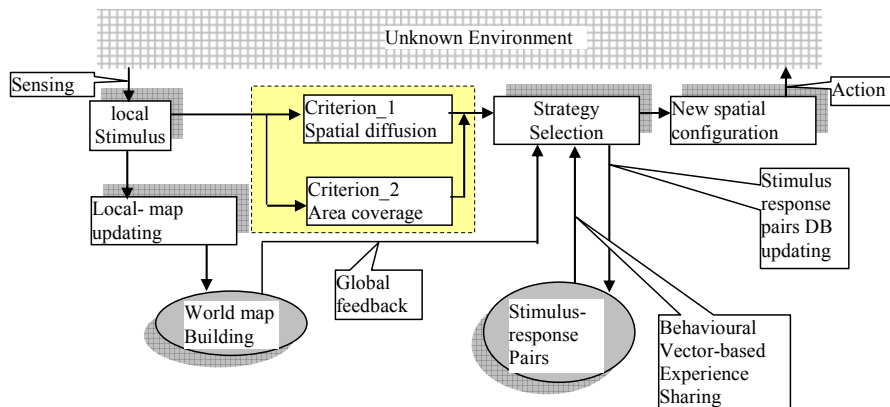


Figure 6

The architecture of the system

Inside the dashed lines the GA module can be seen. For *chromosome representation* let us define a 2D coordinate frame, centred at the current location of the agent, and square bounded, where the sides are determined by the maximal step size ($2 \cdot d_m + l$). In accordance with this, the local region for the agent's next movement is: $x', y' \in [-d_m, d_m]$. Moreover, let us suppose that ($2 \cdot d_m + l$) corresponds to a binary string L , following from the fact that location within the local region can be represented by two binary values, with length L . For the behaviour evolution of a single agent, we can use a chromosome of length $2L$, and for a group with M robots it is $2LM$.

The fitness functions – In this system 3 fitness functions are used, out of which the 1st is the *general* fitness function (f_g), used for exploring less confident regions and for avoiding the repetition of other agents' work [13].

$$f_g = \prod_{i=1}^m \left\{ (1 - \max \{w_i^{tk}\}) \prod_{j=1}^{m_e} \sqrt[4]{d_{ij} - R_1} \right\}; \quad (38)$$

where w_i^{tk} is the confidence weight corresponding to the location of agent i , then m is the number of agents grouped together during one evolutionary movement step, m_e is the number of robots which do not belong to m , that is the inter-distance between two robots i and j is greater than R_1 ($d_{ij} > R_1$). The 2nd and 3rd fitness functions are special functions, and correspond to the criteria of multi-robot spatial diffusion and area coverage, see relations (34), (35).

$$f_1 = \prod_{i=1}^{m_d-1} \prod_{j=i+1}^{m_d} \sqrt{d_{ij} - R_2}; \quad (39)$$

$$f_2 = \frac{\sqrt{\Delta v}}{\prod_{i=1}^{m_c} \xi_i};$$

where, m_d is the number of robots with inter-distances $d_{ij} > R_2$, where m_c is the number of area-covering robots, Δv is the number of location visited by agents m_c and ξ_i is the proximity distance between robot i and other agents. The complete fitness function can be defined as follows:

$$F = \begin{cases} f_g \cdot f_1, & \text{for spatially diffusing robots} \\ f_g \cdot f_2, & \text{for area - covering robots} \end{cases}; \quad (40)$$

8 The System Error

At simulation systems the question of system errors is not avoidable. There are several possibilities to error definition. Usually at models related to mobile robots, we can define errors arising from: *a)* non-ideal mathematical models, *b)* discretization of the work space. Of course each of these errors is repairable. The 1st is repairable by the more exact mathematical models, and the 2nd one by scaling. In the present system, the error arising from discretization of the WS is formally defined as follows [13]:

$$\varepsilon' \cong \sqrt{\frac{1}{K} \sum_{j=1}^K (\tilde{U}_j^t - \hat{U}_j)^2}; \quad (41)$$

where K , denotes the total number of locations in the potential field map, and \tilde{U}_j^t, \hat{U}_j belongs to the estimated and true potential field values at position $P_j(x_j, y_j)$.

Conclusion

It is a simulation system for potential field building process in the multi-agent domain that has been described in this paper. The aim is to provide an opportunity for studying the behaviour vectors of agents, for the sake of selecting the near-optimal behaviour.

The aims stated in the first section (aims & motivation) have been fulfilled. A working multi-agent based simulation model has been created and the features mentioned in the second section (introduction), namely “more efficiency”, have “more or less” been realised as well. Let us look at one of the most important elements of the list: “faster and more accurate”. An unambiguous answer is given in Appendix 5, where on Figure 12 it is clearly seen that in case of a single agent, the system was not able to create the potential field in 30 steps, while in case of 3 or 6 agents (Figures 13 and 14) it was accomplished successfully. Another conception of mine was that the basic problems of trapping and oscillation will be solved by a GA algorithm. The idea has proved to be successful too, as seen in the 5th interval on Figure 15b in Appendix 6, where in case of single agent the problem is clearly visible, while on Figures 16 and 17 (in case of 3 or 6 agents) this problem is not present. The reason why I used the words “more or less” above is because I expected slightly better results from the aspect of area coverage. In my view, in case of 6 agents it is possible to improve the area coverage by tuning the GA parameters and fitness functions.

Appendices

This paper contains 6 appendices. In Appendix 1 the geometrical 2D map of the WS and its exact potential field can be seen. Appendix 2 includes tables with system parameters, GA parameters and computer parameters, used in the simulation. In Appendix 3 the starting positions of mobile agents can be found, while Appendix 4 contains the table and graph of running times with different number of agents and sensors. In Appendix 5 the resulting PFs are seen, also built up by different number of agents and sensors on the agents, while in Appendix 6 the wandered trajectories are represented in 6 intervals.

Appendix 1

The Workspace and its exact potential field

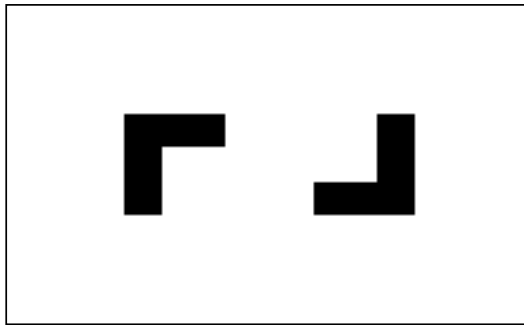


Figure 7
The *WS* in 2D with two obstacles

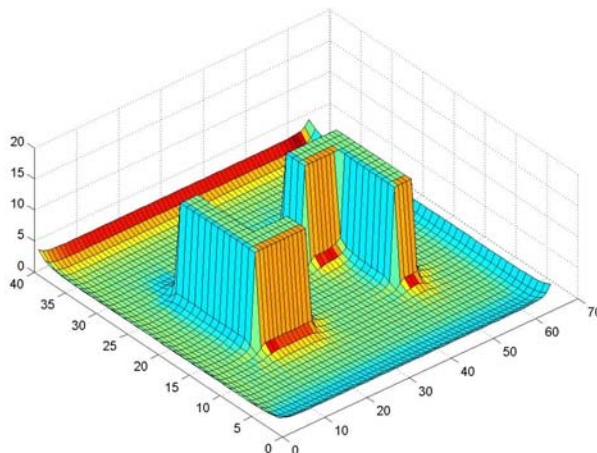


Figure 8
The exact *PF* of *WS*

Appendix 2

Tables of *parameters* used in the system and in the *GA*

Table 1
System parameters

Description	Unit	Value
The loaded <i>WS</i> size	unit	195 x 120
<i>WS</i> resolution (grid width)	grid	3
Normalised <i>WS</i>	$\frac{X}{GridWidth}; \frac{Y}{GridWidth}$	40 x 65
Maximum movement step (d_m)	location	7
Behaviour-vector increment		0,2
Threshold distance <i>R1</i>	grid	10
Threshold distance <i>R2</i>	grid	15

Table 2
Parameters used in *GA*

Description	Unit	Value
Robot description (in <i>GA</i>)	bit	8
Population size (<i>P</i>)		20/3045/65/90/120
Generations per step		8/12/18/26/36/48
Crossover probability (p_c)		0,6
Mutation probability (p_m)		0,1/0,05/0,005

Table 3
Computer parameters

Description	Unit	Value
Operation system		WIN-XP, prof.
Processor clock	GHz	1,60
RAM size	Mbyte	512

Appendix 3

The starting positions of the mobile robots

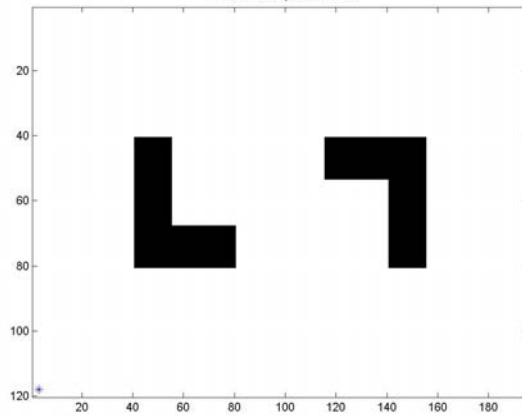


Figure 9
Starting position of 1 agent

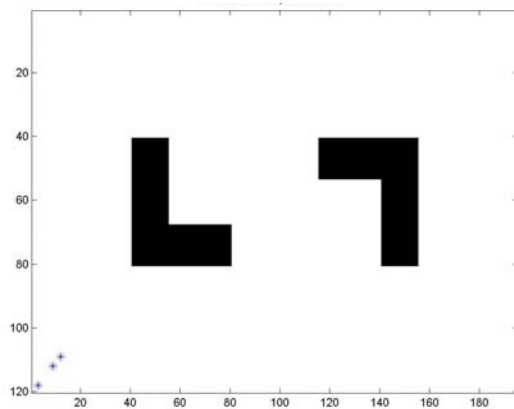


Figure 10
Starting positions of 3 agents

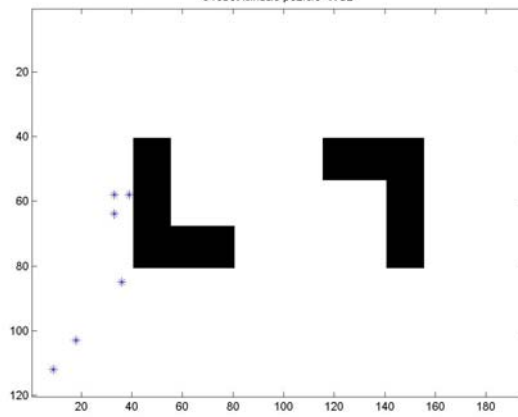



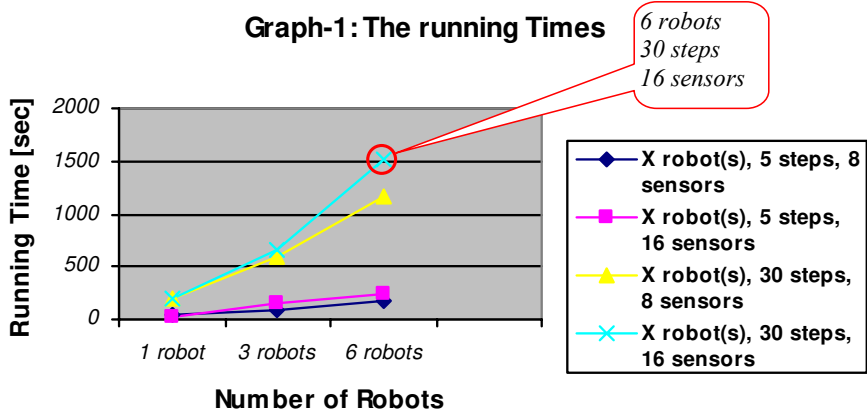
Figure 11
Starting positions of 6 agents

Appendix 4

Table of results

Table 4
The Running Times

 ws2_2d.jpg	<i>Running Times</i> (in seconds)			
	5		30	
Maximum number of steps	8	16	8	16
Number of sensors				
Number of robots ↓				
1	37	31	208	195
3	77	150	595	657
6	177	240	1162	1521



Appendix 5

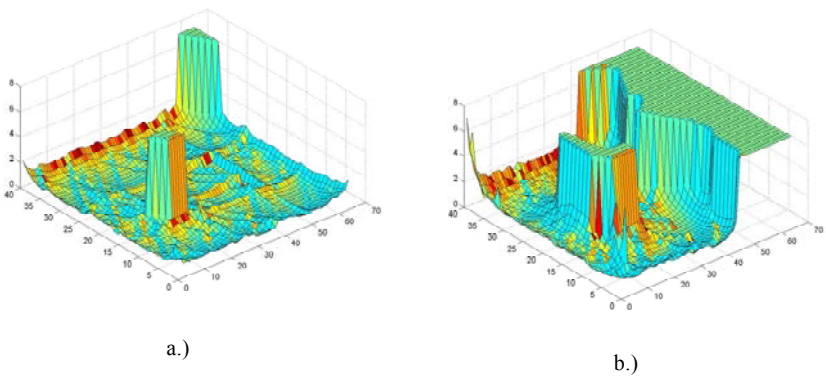


Figure 12
The resulting *PF* of *WS*, built up by *1* agent in *30* steps
a) 8 sensors, b) 16 sensors

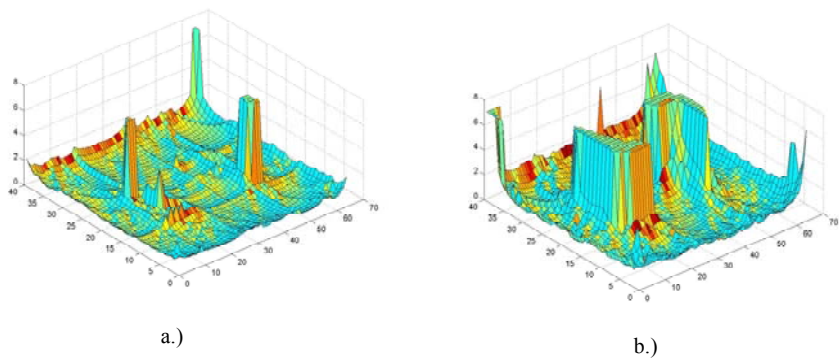


Figure 13
The resulting *PF* of *WS*, built up by *3* agents in *30* steps:
a) 8 sensors, b) 16 sensors

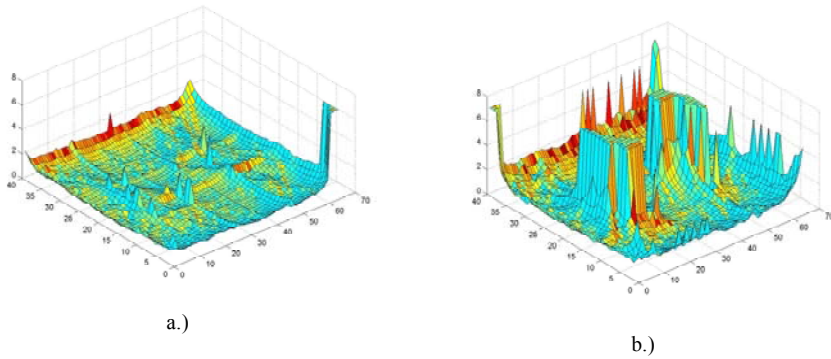


Figure 14
The resulting *PF* of *WS* built up with 6 agents in 30 steps:
a) 8 sensors, b) 16 sensors

Appendix 6

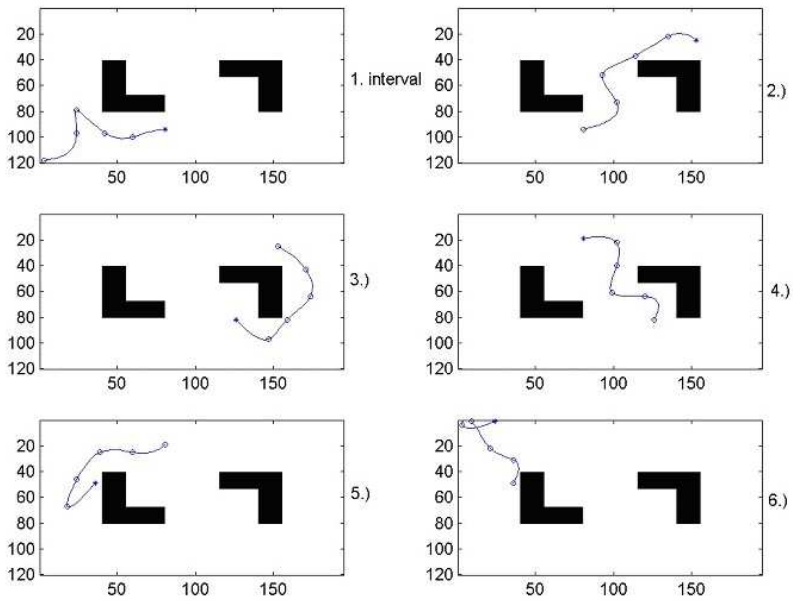


Figure 15a
Trajectories of 1 agent, in 6 intervals, MaxRunStep=30, 8 sensors

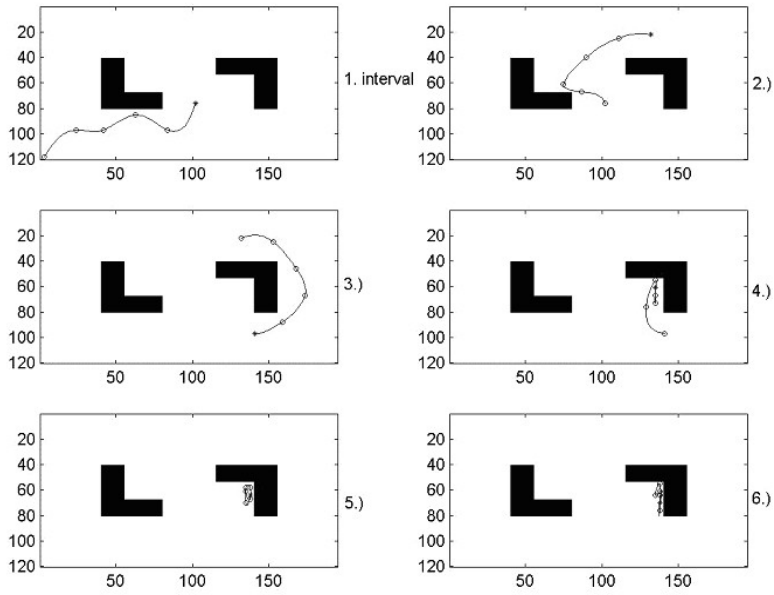


Figure 15b

Trajectories of 1 agent, in 6 intervals, MaxRunStep=30, 16 sensors

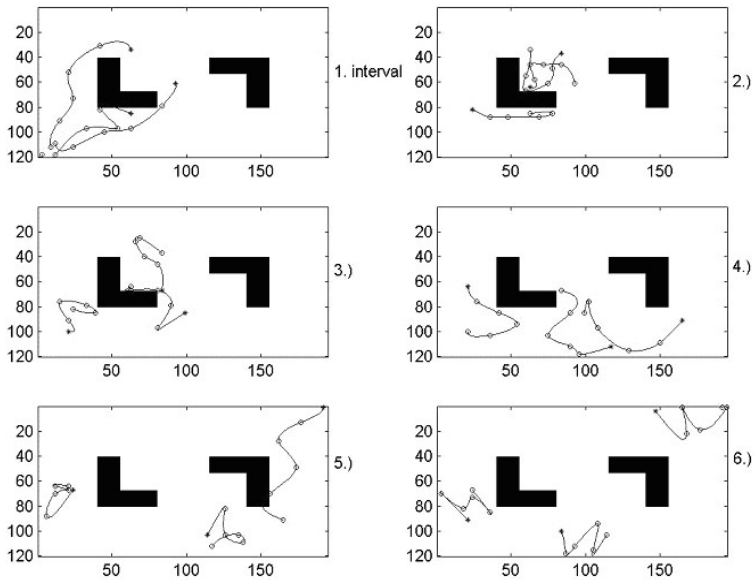


Figure 16a

Trajectories of 3 agents, in 6 intervals, MaxRunStep=30, 8 sensors

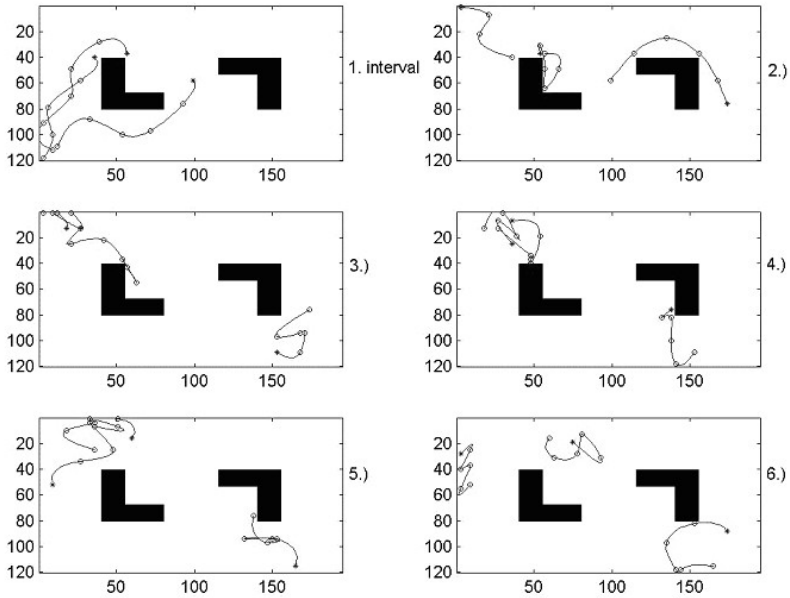


Figure 16b

Trajectories of 3 agents, in 6 intervals, MaxRunStep=30, 16 sensors

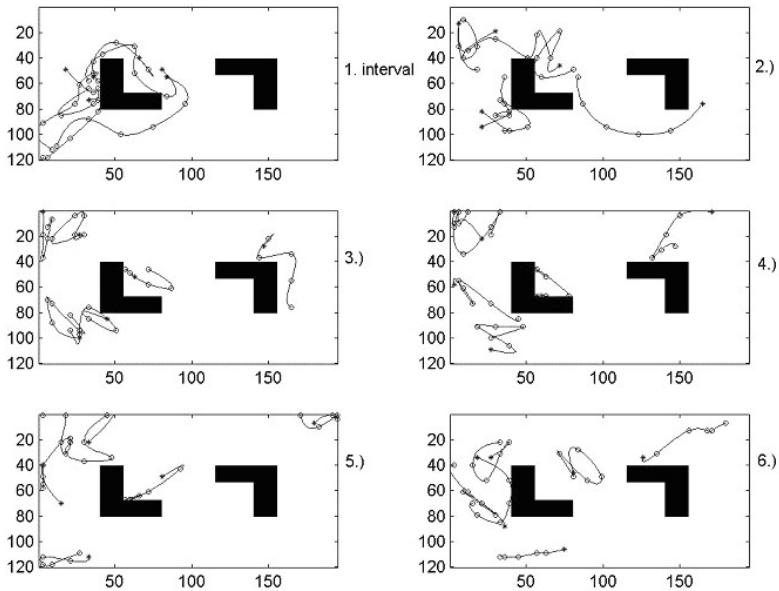


Figure 17a

Trajectories of 6 agents, in 6 intervals, MaxRunStep=30, 8 sensors

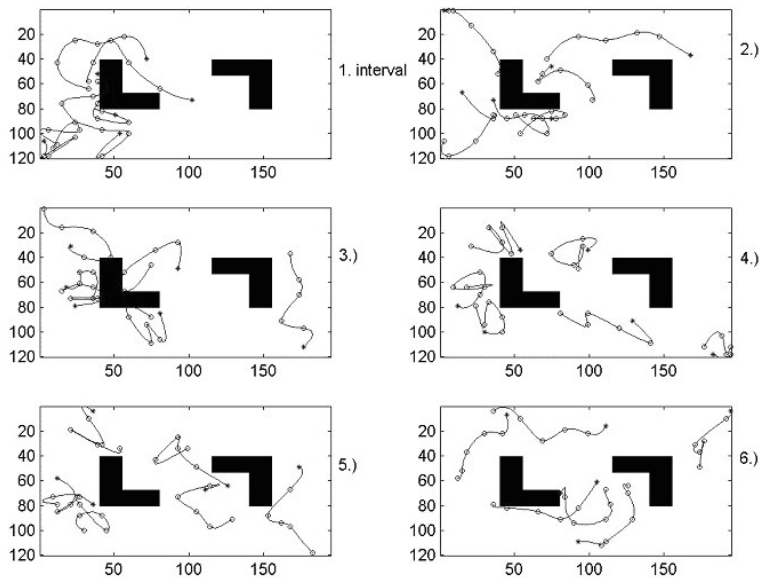


Figure 17b

Trajectories of 6 agents, in 6 intervals, MaxRunStep=30, 16 sensors

References

- [1] S. Mizik, P. Baranyi, P. Korondi, M. Sugiyama: *Virtual Training of Vector Function-based Guiding Styles*; Transactions on Automatic Control and Computer Science, ISSN 1224/600X Vol. 46(60) No. 1, pp. 81-86. 2001
- [2] J. Vaščák: *Navigation of Mobile Robots Using Potential Fields and Computational Intelligence Means*; Acta Polytechnica Hungarica, ISSN 1785-8860, Vol. 4, No. 1, pp. 63-74, 2007
- [3] I. Nagy, A. L. Bencsik: *A Simulation System for Behaviour-based Potential Field Building in Multi-Agent Mobile Robot System*; Proc. of the 3rd IAESTED International Conference on Computational Intelligence, pp. 7-12, ISBN: 978-0-88986-672-0, Canada, 2007
- [4] T. Fukuda, S. Nakagawa: *A Dynamically Reconfigurable Robotic System*; In Proc. of the International Conference on Industrial Electronics, Control and Instrumentation, pp. 588-595, Cambridge, MA, 1987
- [5] H. Asama, A. Matsumoto, Y. Ishida: *Design of an Autonomous and Distributed Robot System: ACTRESS*; In proc. of the IEEE/RSJ, International Workshop on Intelligent Robots and Systems, pp. 283-290, Tsukuba, 1989

- [6] D. Goldberg, M. J. Mataric: *Coordinating Mobile Robot Group Behaviour Using a Model of Interaction Dynamics*; Proc. of the 3rd Int. Conf. on Autonomous Agents, pp. 100-107, Seattle, 1999
- [7] P. Pirjanian, M. J. Mataric: *Multi-Robot Target Acquisition Using Multiple Objective Behaviour Coordination*; In proc. of the IEEE International Conference on Robotics and Automation, San Francisco, 2000
- [8] M. J. Mataric: *Behaviour-based Robotics as a Tool for Synthesis of Artificial Behaviour and Analysis of Natural Behaviour*; Trends in Cognitive Science, 2(3), pp. 82-87, 1998
- [9] Gh. Tecuci: *Building Intelligent Agents*; Academic Press, San Diego, Cal., 1998
- [10] J. Kelemen: *Agents from Functional-Computational Perspective*; Acta Polytechnica Hungarica, ISSN 1785-8860, Vol. 3, No. 4, pp. 37-54, 2006
- [11] B. Frankovič, T-T. Dang, I. Budinská: *Agents' Coalitions Based on a Dynamic Programming Approach*; Acta Polytechnica Hungarica, ISSN 1785-8860, Vol. 5, No. 2, pp. 5-21, 2008
- [12] T. Fukuda, G. Iritani: *Construction Mechanism of Group Behaviour with Cooperation*; In Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 535-542, Pittsburgh, 1995
- [13] J. Liu, J. Wu: *Multi-Agent Robotic Systems*; CRC Press LLC, ISBN 0-8493-2288-X, 2001
- [14] G. Dudek, M. Jenkin: *Computational Principles of Mobile Robotics*; Cambridge University Press, ISBN 0 521 56021 7, 2000
- [15] I. Nagy: *Genetic Algorithms Applied for Potential Field Building in Multi-Agent Robotic System*; In Proc. of the IEEE International Conference on Computational Cybernetics, ICC 2003, pp. 105-108, Siófok, Hungary, 2003

Control of a Uniform Step Asymmetrical 9-Level Inverter Based on Artificial Neural Network Strategy

Rachid Taleb¹, Abdelkader Meroufel², Patrice Wira³

¹Electrical Engineering Department, Hassiba Ben Bouali University
BP 151 Hay Es-Salam Chlef, Algeria, e-mail: murad72000@yahoo.fr

²Intelligent Control and Electrical Power Systems Laboratory (ICEPS)
Djillali Liabes University, BP 89 Sidi Bel-Abbes, Algeria

³Laboratoire Modélisation, Intelligence, Processus et Systèmes (MIPS)
Université de Haute Alsace, 68093 Mulhouse, France

Abstract: A neural implementation of a harmonic elimination strategy for the control a uniform step asymmetrical 9-level inverter is proposed and described in this paper. A Multi-Layer Perceptrons (MLP) neural network is used to approximate the mapping between the modulation rate and the required switching angles. After learning, the neural network generates the appropriate switching angles for the inverter. This leads to a low-computational-cost neural controller which is therefore well suited for real-time applications. This neural approach is compared to the well-known Multi-Carrier Pulse-Width Modulation (MCPWM). Simulation results demonstrate the technical advantages of the neural implementation of the harmonic elimination strategy over the conventional method for the control of an uniform step asymmetrical 9-level inverter. The approach is used to supply an asynchronous machine and results show that the neural method ensures a highest quality torque by efficiently canceling the harmonics generated by the inverter.

Keywords: Uniform step asymmetrical multilevel inverter, Harmonics Elimination Strategy, Artificial Neural Networks, Multi-Layer Perceptron, Multi-Carrier Pulse-Width Modulation

1 Introduction

Inverters are widely used in modern power grids; a great focus is therefore made in different research fields in order to develop their performance. Three-level inverters are now conventional apparatus but other topologies have been attempted this last decade for different kinds of applications [1]. Among them, Neutral Point Clamped (NPC) inverters, flying capacitors inverters also called imbricated cells, and series connected cells inverters called cascaded inverters [2].

This paper is a study about a three-phase multilevel converter based on series connected single phase inverters (partial cells) in each phase. A multilevel converter with k partial inverters connected in serial is presented by Fig. 1. In this configuration, each cell of rank $j = 1, \dots, k$ is supplied by a dc-voltage source u_{dj} . It has been shown that feeding partial cells with unequal dc-voltages (asymmetric feeding) increases the number of levels of the generated output voltage without any supplemental complexity to the existing topology [3]. These inverters are referred to as "Asymmetrical Multilevel Inverters" or AMI.

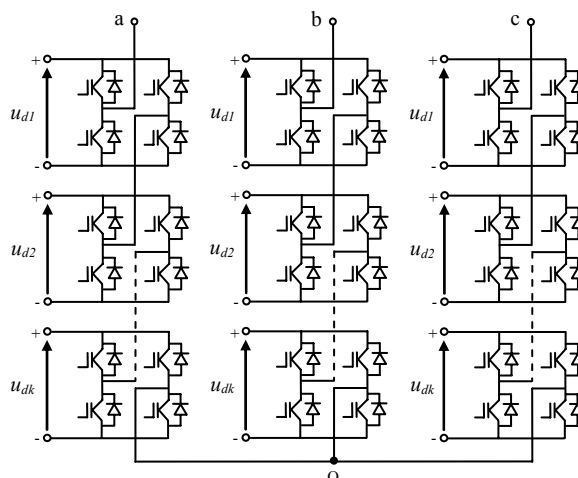


Figure 1

Three-phase structure of a multilevel converter with k partial monophased inverters series connected per phase

Some applications such as active power filtering need inverters with high performances [4]. These performances are obtained if there are still any harmonics at the output voltages and currents. Different Pulse-Width Modulation (PWM) control-techniques have been proposed in order to reduce the residual harmonics at the output and to increase the performances of the inverters [5]. The most popular one is probably the multi-carrier PWM technique [6] which shifts the harmonics to high frequencies by using high-frequency carriers. However, electronic devices and components have limited switching-frequencies. High-frequency carriers are therefore limited by this constraint. An alternative solution consists in adapting the principle of the Harmonics Elimination Strategy (HES) to AMIs [7]. The HES allows canceling the critical harmonic distortions and therefore controlling the fundamental component of the signal by using electronic devices with low switching frequencies.

The principle of this technique relies on the resolution of a system of non linear equations to elaborate the switching angle control signals for the electronic devices [8]. Practically, the implementation of this method requires memorizing

all the firing angles which is complex and needs considerable computational costs. Mathematical solutions with limited computational costs are therefore preferably used for real-time applications. The approach can be achieved with Artificial Neural Networks (ANNs) which are known as parsimonious universal approximators. Their learning from examples leads to robust generalization capabilities [9].

This paper proposes a HES based on ANNs to control a 9-level Uniform Step Asymmetrical Multilevel Inverter (USAMI). The work presented in [10] has been applied to an 11-level USAMI. In this paper, the neural implementation of the HES is applied to 9-level USAMI. Standard Multi-Layer Perceptrons (MLP) [11] are used for approximating the relationship between the modulation rate and the inverters switching angles. The performance of this neural approach is evaluated and compared to the MCPWM technique. The proposed neural strategy is also evaluated when the inverter supplies an asynchronous machine. In this application, it is important that the implemented controller computes appropriate switching angles for the inverters in order to minimize the harmonics absorbed by the asynchronous machine. Performances were successfully achieved, the neural controller demonstrates a satisfying behavior and a good robustness.

The paper is organized as follows. USAMIs are described and modeled in Section 2. Section 3 briefly introduces the well-known MCPWM and brings out the original HES based on a MLP. Section 4 evaluates the proposed neural strategy in computing optimal angles of an inverter used to supply an asynchronous machine. The results show that the neural method cancels the harmonics distortions and supplies the machine with a well-formed sinusoidal voltage waveform. Finally, a summary of the results is presented in the Conclusion.

2 Uniform Step Asymmetrical Multilevel Inverters

Multilevel inverters generate at the ac-terminal several voltage levels as close as possible to the input signal. Fig. 2 for example illustrates the N voltage levels u_{s1} , u_{s2} , ... u_{sN} composing a typical sinusoidal output voltage waveform. The output voltage step is defined by the difference between two consecutive voltages. A multilevel converter has a uniform or regular voltage step, if the steps Δu between all voltage levels are equal. In this case the step is equal to the smallest dc-voltage, u_{d1} [6]. This can be expressed by

$$u_{s2} - u_{s1} = u_{s3} - u_{s2} = \dots = u_{sN} - u_{s(N-1)} = \Delta u = u_{d1} \quad (1)$$

If this is not the case, the converter is called a non uniform step AMI or irregular AMI. An USAMI is based on dc-voltage sources to supply the partial cells (inverters) composing its topology which respects to the following conditions [6]:

$$\begin{cases} u_{d1} \leq u_{d2} \leq \dots \leq u_{dk} \\ u_{dj} \leq 1 + 2 \sum_{l=1}^{j-1} u_{dl} \end{cases} \quad (2)$$

where k represents the number of partial cells per phase and $j = 1, \dots, k$. The number of levels of the output voltage can be deduced from

$$N = 1 + 2 \sum_{j=1}^k u_{dj} \quad (3)$$

This relationship fundamentally modifies the number of levels generated by the multilevel topology. Indeed, the value of N depends on the number of cells per phase and the corresponding supplying dc-voltages.

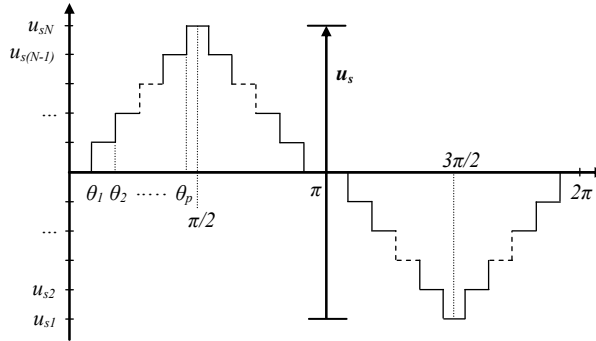


Figure 2

Typical output voltage waveform of a multilevel inverter

Equation (3) accepts different solutions. With $k = 3$ for example, there are two possible combinations of supply voltages for the partial inverters in order to generate a 11-level global output, i.e., $(u_{d1}, u_{d2}, u_{d3}) \in \{(1, 1, 3); (1, 2, 2)\}$, and there are three possible combinations to generate a 15-level global output, i.e., $(u_{d1}, u_{d2}, u_{d3}) \in \{(1, 1, 5); (1, 2, 4); (1, 3, 3)\}$. Fig. 3 shows the possible output voltages of the three partial cells of the 9-level inverter with $k = 3$. The dc-voltages of the three cells are $u_{d1} = 1p.u.$, $u_{d2} = 1p.u.$ and $u_{d3} = 2p.u.$. The output voltages of each partial inverter are noted u_{p1} , u_{p2} and u_{p3} and can take three different values: $u_{p1} \in \{-1, 0, 1\}$, $u_{p2} \in \{-1, 0, 1\}$ and $u_{p3} \in \{-2, 0, 2\}$. The result is a generated output voltage with 9 levels: $u_s \in \{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$. Some levels of the output voltage can be generated by different commutation sequences. For example, there are four possible commutation sequences resulting in $u_s = 2p.u.$: $(u_{p1}, u_{p2}, u_{p3}) \in \{(-1, 1, 2); (0, 0, 2); (1, -1, 2); (1, 1, 0)\}$. These redundant combinations can be selected in order to optimize the switching process of the inverter [12].

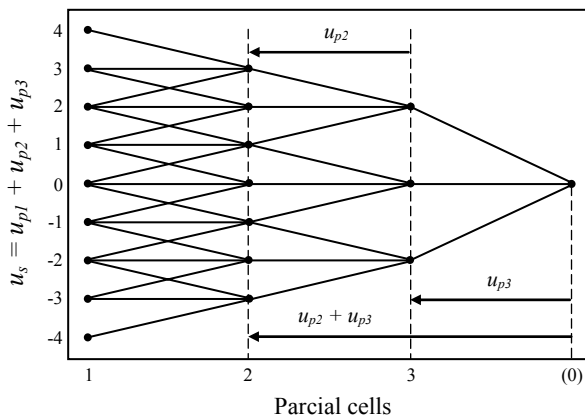


Figure 3

Possible output voltages of each partial inverter to generate $N=9$ levels with $k=3$ cells per phase
(with $u_{d1} = 1p.u.$, $u_{d2} = 1p.u.$ and $u_{d3} = 2p.u.$)

These different possibilities offered by the output voltage of the partial inverters, and the redundancies among them to deliver a same output voltage level, can be considered as degrees of freedom which can be exploited in order to optimize the use of a AMI.

3 Multilevel Inverters Control Strategies

Several modulation strategies have been proposed for symmetrical multilevel converters. They are generally derived from the classical modulation techniques used for more traditional converters [12]. Among these methods, the most common used is the multi-carrier sub-harmonic PWM technique. This modulation method can also be used to control asymmetrical multilevel power converters. In the case of AMIs, other kinds of modulation can be used [3].

In this Section, we briefly introduce the MCPWM technique. We also propose a HES based on ANNs. These control strategies will be compared by computer simulations. The objective is to elaborate optimized switching angles for an 9-level USAMI. The inverter is then employed to supply an asynchronous machine.

3.1 Multi-Carrier PWM (MCPWM)

The principle of the MCPWM is based on a comparison of a sinusoidal reference waveform with vertically shifted carrier waveforms. $N-1$ carriers are required to generate N levels. As shown in Fig. 4, the carriers are in continuous bands around the reference zero. They have the same amplitude A_c and the same frequency f_c . The sine reference waveform has a frequency f_r and an amplitude A_r . At each

instant, the result of the comparison is 1 if the triangular carrier is greater than the reference signal and 0 otherwise. The output of the modulator is the sum of the different comparisons which represents the voltage level. The strategy is therefore characterized by the two following parameters [6], respectively called the modulation index and the modulation rate:

$$m = \frac{f_c}{f_r} \tag{4}$$

$$r = \frac{2}{N-1} \frac{A_r}{A_c} \tag{5}$$

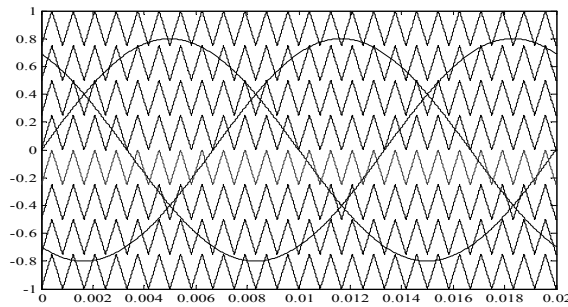


Figure 4

Multi-carrier PWM generation for $N = 9$ levels (with $m = 24$ and $r = 0.8$)

We propose to develop a 9-level inverter composed of $k = 3$ partial inverters per phase with the following dc-voltage sources: $u_{d1} = 1p.u.$, $u_{d2} = 1p.u.$ and $u_{d3} = 2p.u.$. The output voltage V_{ab} and its frequency representation are respectively presented by Fig. 5 and Fig. 6. The output voltages u_{p1} , u_{p2} and u_{p3} of each partial inverter and the resulting voltage for the first phase are represented by Fig. 7.

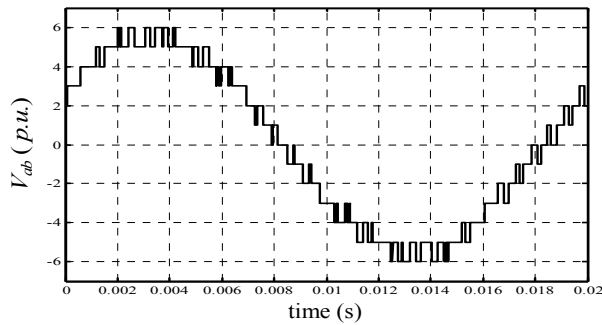


Figure 5

Output voltage V_{ab} of the 9-level USAMI controlled by the MCPWM (with $m = 24$ and $r = 0.8$)

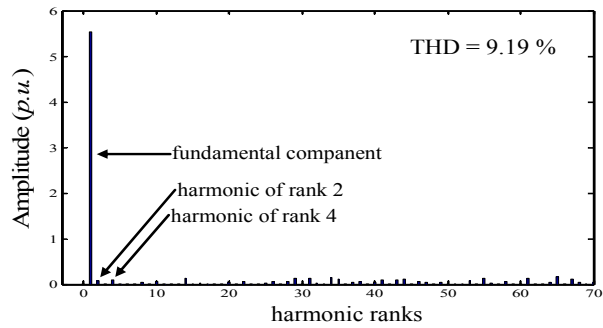


Figure 6

Frequency content of the output voltage V_{ab} with the MCPWM strategy (with $m = 24$ and $r = 0.8$)

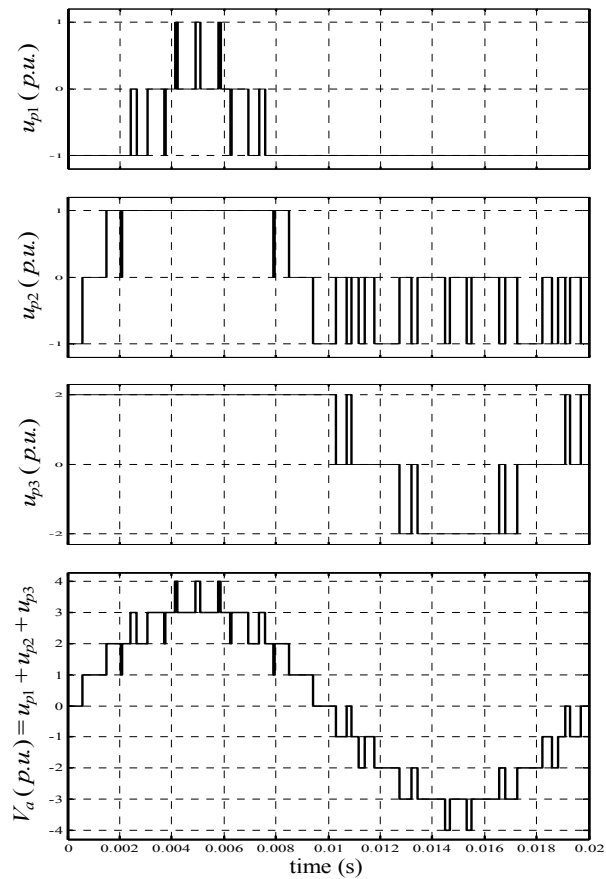


Figure 7

Output voltages of each partial inverter and total output voltage V_a of the 9-level USAMI controlled by the MCPWM (with $m = 24$ and $r = 0.8$)

3.2 Harmonics Elimination Strategy with ANNs

A) *Harmonics Elimination Strategy (HES)*: The HES is based on the Fourier analysis of the generated voltage u_s at the output of the USAMI (see Fig. 2) [13]. This voltage is symmetric in a half and a quarter of a period. As a result, the even harmonic components are null. The Fourier series expansion for the u_s voltage is thus:

$$\begin{cases} u_s = \sum_{n=1}^{\infty} u_n \sin n\omega t \\ u_n = \frac{4u_{d1}}{n\pi} \sum_{i=1}^p \cos n\theta_i \end{cases} \quad (6)$$

where u_n represents the amplitude of the harmonic term of rank n , $p = (N - 1)/2$ is the number of switching over a quarter of a period, and θ_i are the switching angles ($i = 1, 2, \dots, p$).

The p switching angles in (6) are calculated by fixing the amplitude of the fundamental term and by canceling the $p - 1$ other harmonic terms. Practically, four switching angles ($\theta_1, \theta_2, \dots, \theta_4$) are necessary for canceling the three first harmonics terms (i.e., harmonics with a odd rank and non multiple of 3, therefore 5, 7 and 11) in the case of a three phase 9-level USAMI composed of $k = 3$ partial inverters per phase supplied by the dc-voltages $u_{d1} = 1p.u.$, $u_{d2} = 1p.u.$ and $u_{d3} = 2p.u.$ These switching angles can be determined by solving the following system of non linear equations:

$$\begin{cases} \sum_{i=1}^4 \cos \theta_i = \pi r \\ \sum_{i=1}^4 \cos n\theta_i = 0 \text{ for } n \in \{5, 7, 11\} \end{cases} \quad (7)$$

where $r = u_1/4u_{d1}$ is the modulation rate. The solution of (7) must also satisfy

$$\theta_1 < \theta_2 < \theta_3 < \theta_4 < \frac{\pi}{2} \quad (8)$$

and can be solved by applying the Newton-Raphson method. This method returns all the possible combinations of the switching angles for different values of r . The result is represented by Fig. 8 where one can see the presence of two possible solutions of angles for $0.70 \leq r \leq 0.76$. On the other side, the system does not accept any solution for $r < 0.629$, $0.64 < r < 0.7$ and $0.897 < r < 0.921$. The system has an unique solution for all the other values of r .

In the case of two possible solutions for an angle θ_i , the criteria for selecting one of them can be the Total Harmonic Distortion (THD). The best angle values are therefore the ones leading to the lowest THD. The THD is a quantifiable expression for determining how much the signal has been distorted. The greater

are the amplitudes of the harmonics, the greater are the distortions. The THD is defined by:

$$THD = \sqrt{\sum_{n=2}^{\infty} \left(\frac{1}{n} \sum_{i=1}^{p=4} \cos n\theta_i \right)^2} / \sum_{i=1}^{p=4} \cos \theta_i \tag{9}$$

This is shown in Fig. 9 corresponding to the solutions given in Fig. 8. Choosing the switching angles based on this criteria, the multiple switching angle solutions given in Fig. 8 reduce to the single set of solutions given in Fig. 10, and the corresponding THD is shown in Fig. 11.

The control of an AMI with the HES in a real-time application requires to memorize all the switching angles. A considerable computational memory space must therefore be involved for the implementation of this control law.

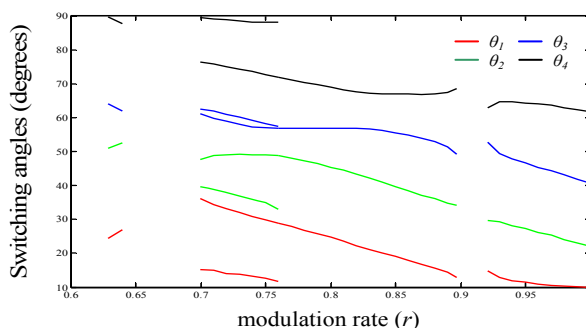


Figure 8
All switching angles versus r

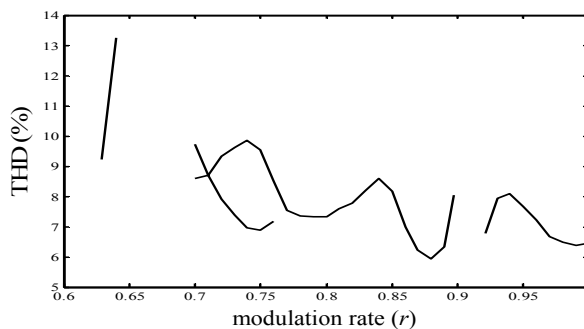


Figure 9
THD versus r for all switching angles

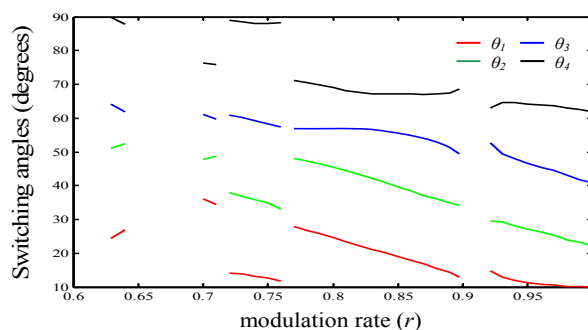


Figure 10

Switching angles versus r leading to the lowest THD

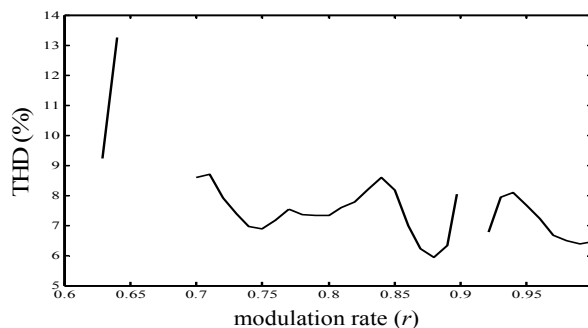


Figure 11

THD versus r for the switching angles that result in the lowest total harmonic distortion

B) Application of ANNs: ANNs have gained increasing popularity and have demonstrated superior results compared to alternative methods in many studies. Indeed, ANNs are able to map underlying relationship between input and output data without prior understanding of the process under investigation. This mapping is achieved by adjusting their internal parameters called weights from data. This process is called the learning or the training process. Their interest comes also from their generalization capabilities, i.e., their ability to deliver estimated responses to inputs that were not seen during training. Hence, the application of ANNs to complex relationships and processes makes them highly attractive for different types of modern problems [9, 14].

We use a neural network to learn the switching angles previously provided by the Newton-Raphson method. The approach aims to replace the painful memorization of the angles in order to make its implementation realizable in a real-time application. MLPs [11] are well suited for this task. Associated to the backpropagation learning rule, they are known as universal approximators [9].

An MLP network is composed of a number of identical units called neurons organized in layers, with those on one layer connected to those on the next layer

(except for the last layer or output layer). Indeed, MLPs architecture is structured into an input layer of neurons, one or more hidden layers and one output layer. Neurons belonging to adjacent layers are usually fully connected and the activation function of the neurons is generally sigmoidal or linear. In fact, the various types and architectures are identified both by the different topologies adopted for the connections and by the choice of the activation function.

Some parameters of ANNs can not be determined from an analytical analysis of the process under investigation. This is the case of the number of hidden layers and the number of neurons belonging to them. Consequently, they have to be determined experimentally according to the precision which is desired for the estimation. The number of inputs and outputs depends from the considered process. In our application, the MLP has to map the underlying relationship between the modulation rate (input) and the p switching angles (output). The MLP shown in Fig. 12 is composed by one input neuron and p output neurons.

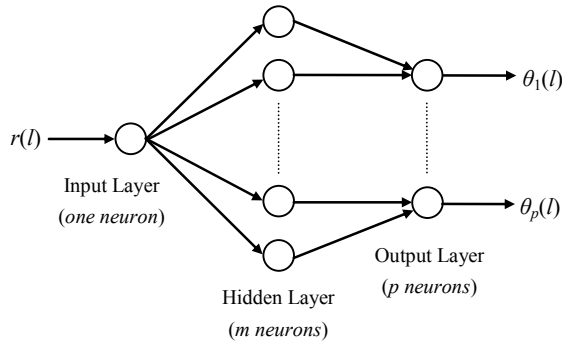


Figure 12

Multi-Layer Perceptron network topology ($1 \times m \times p$) used to generate switching angles

The MLP must be trained in order to adjust and to find the adequate weights. This is achieved by using probabilistic learning techniques and with data from the process under investigation. The training data consists of the inputs R and the corresponding desired output vectors S :

$$R = [r(1) \dots r(l) \dots r(n)] \quad (10)$$

$$S = [S(1) \dots S(l) \dots S(n)] = \begin{bmatrix} \theta_1(1) & \dots & \theta_1(l) & \dots & \theta_1(n) \\ \vdots & & \vdots & & \vdots \\ \theta_p(1) & \dots & \theta_p(l) & \dots & \theta_p(n) \end{bmatrix} \quad (11)$$

In the last two expressions, $l = 1 \dots n$, where n is the number of examples. For a given input $r(l)$, the MLP computes an estimated output vector $\hat{S}(l) = [\hat{\theta}_1(l) \dots \hat{\theta}_p(l)]^T$ that must be as close as possible to the ideal desired output $S(l)$. The difference $E(l) = (\hat{S}(l) - S(l))^2$ constitutes the squared output error for example l that is used by the training algorithm to correct the weights of the

neurons. This is repeated for the n samples composing the training data set until convergence is reached. The learning is achieved with the backpropagation algorithm [9].

After the training process, the MLP is able to estimate the angles corresponding to an input $r(l)$. In other words, the MLP has learned the functions $\theta_i = f_i(r)$ with $i = 1 \dots 4$. By approximating these functions, the MLP will be able to deliver the angles for the real-time control of the inverter.

4. Results

4.1 Learning Performance

The most significant harmonics in power systems are those of rank 5, 7 and 11. These harmonics are essentially present at the output of the USAM. We therefore chose to use explicitly cancel them with the MLP-based approach. The learning is elaborated with the Newton-Raphson method and the optimal angles are the ones resulting in the lowest THD when several solutions exist.

A MLP with one hidden layer is used with a training set composed of $n = 33$ examples. The MLP takes one input, i.e., the modulation rate, and delivers four outputs which are the switching angles. Several tests have been conducted for determining the number of neurons of the hidden layer. These tests have been achieved because there are no generally acceptable theories in the literature for choosing the number of hidden layers and hidden neurons for a specific application. The number of neurons in the hidden layers is important in the sense that it affects the learning convergence and overall generalization property of the MLP.

Table 1
Approximation errors with various sizes of the MLP

Number of neurons in the MLP layers	Required iterations	Learning error (degrees)
$1 \times 2 \times 4$	10 000	4.372
$1 \times 4 \times 4$	10 000	3.213
$1 \times 6 \times 4$	10 000	2.105
$1 \times 7 \times 4$	10 000	1.467
$1 \times 8 \times 4$	8 943	0.893
$1 \times 9 \times 4$	6 671	0.677
$1 \times 10 \times 4$	4 249	$8 \cdot 10^{-2}$
$1 \times 11 \times 4$	2 637	$4 \cdot 10^{-3}$
$1 \times 12 \times 4$	1 528	$1 \cdot 10^{-3}$

Results are provided by Table 1. According to this table, a MLP with 12 neurons in the hidden layer has been adopted. The approximating error remains the same with 12 neurons as with a higher number of neurons in the hidden layer. This configuration has been chosen after different experiments, it represents the best compromise between computational costs and performances. The other parameters of the MLP are detailed in Table 2.

Table 2
Properties of the MLP

MLP parameters	values
Network configuration	$1 \times 12 \times 4$
Transfer functions	tansig, purelin
Training technique	Levenberg-Marquardt
Learning rate	0.1
Momentum constant	0.9
Training goal	0.001
Training patterns	33
Epochs	1528
Maximum epochs	10 000

The learning convergence of the $1 \times 12 \times 4$ -MLP is reached after 1528 epochs, and leads to angle errors less than 0.001 degrees. The outputs delivered by the MLP are therefore very close to the angles given by the Newton-Raphson method. The estimated angles are represented by Fig. 13. The evolution of the training Sum-Squared-Error (SSE) is shown by Fig. 14.

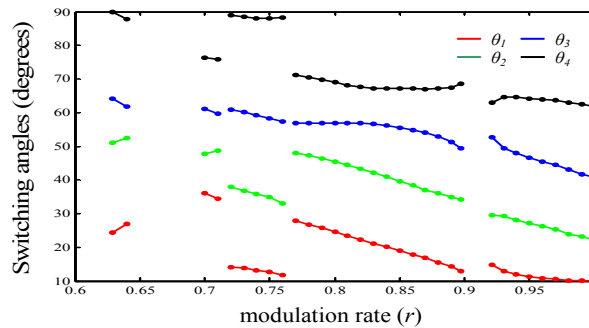


Figure 13

Switching angles versus r estimated by the MLP (•) and calculated with Newton-Raphson method (–)

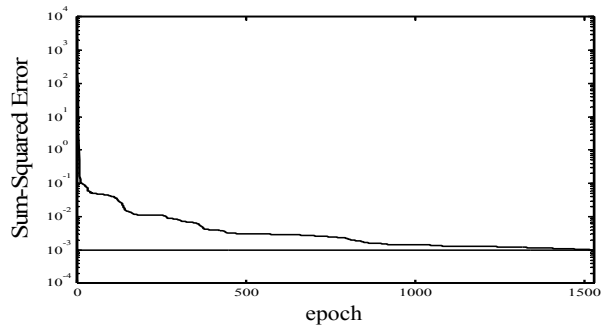


Figure 14
Training error

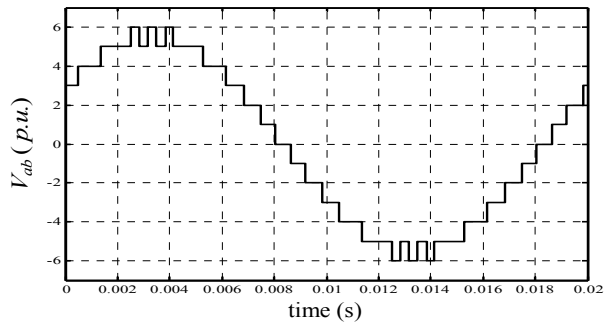


Figure 15

Output voltage V_{ab} of the 9-level USAMI controlled by the proposed neural HES (with $r(l) = 0.8$)

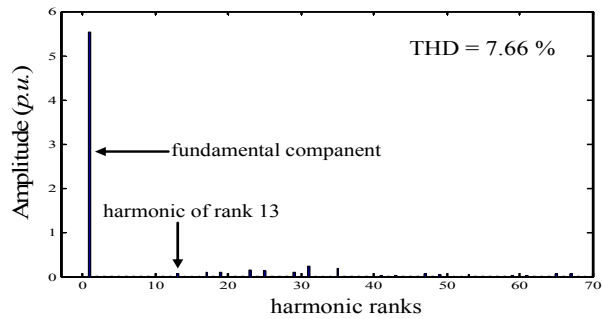


Figure 16

Frequency content of the output voltage V_{ab} with the proposed neural HES (with $r(l) = 0.8$)

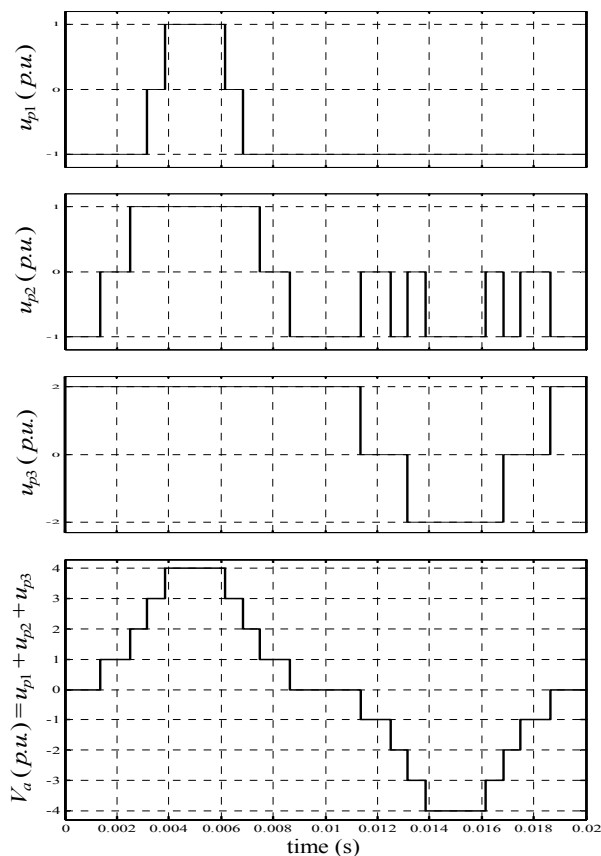


Figure 17

Output voltages of each partial inverter and total output voltage V_a of the 9-level USAMI controlled by the proposed neural HES (with $r(l) = 0.8$)

After learning, the MLP is also able to deliver the angles for inputs which were not present in the training set. These generalization capabilities are very interesting in our application, the neural controller is therefore always able to deliver the control signals for the inverter. For example, Fig. 15 to 17 shows the results obtained for an input which was not in the training set, $r(l) = 0.8$ which theoretically corresponds to $\theta_1 = 24.6999^\circ$, $\theta_2 = 45.5307^\circ$, $\theta_3 = 57.0398^\circ$ and $\theta_4 = 68.8887^\circ$.

4.2 Performance in Supplying an Asynchronous Machine

In order to evaluate the performance and the robustness of the proposed approach, a 9-level USAMI is used to supply an asynchronous machine (with parameters given in the Appendix section). The neural HES is compared to the MCPWM

strategy in controlling the 9-level USAMI. The objective is to use the proposed neural strategy in order to minimize the harmonics absorbed by the asynchronous machine.

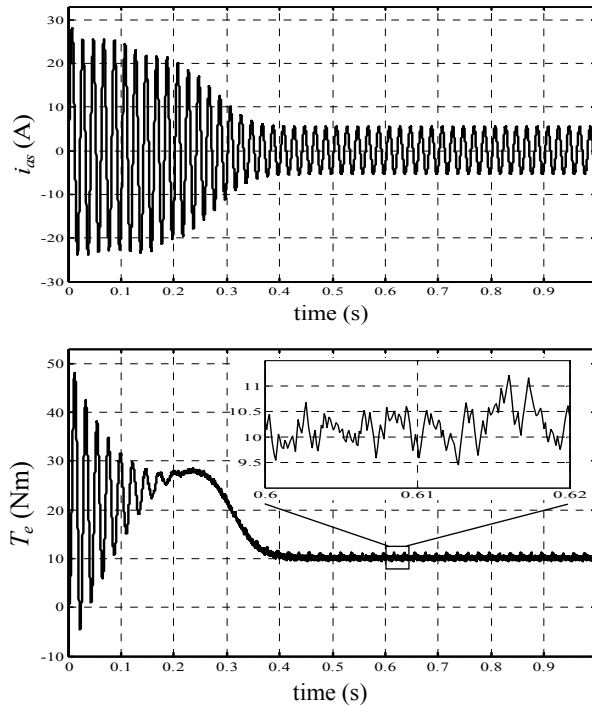


Figure 18

Stator current (top) and electromagnetic torque (bottom) of the asynchronous machine fed by a 9-level USAMI controlled by the MCPWM

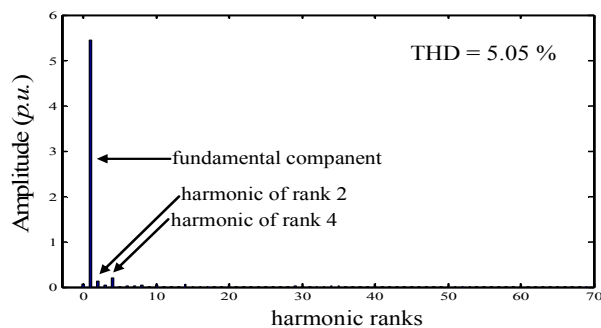


Figure 19

Frequency content of the stator current of the asynchronous machine fed by a 9-level USAMI controlled by the MCPWM

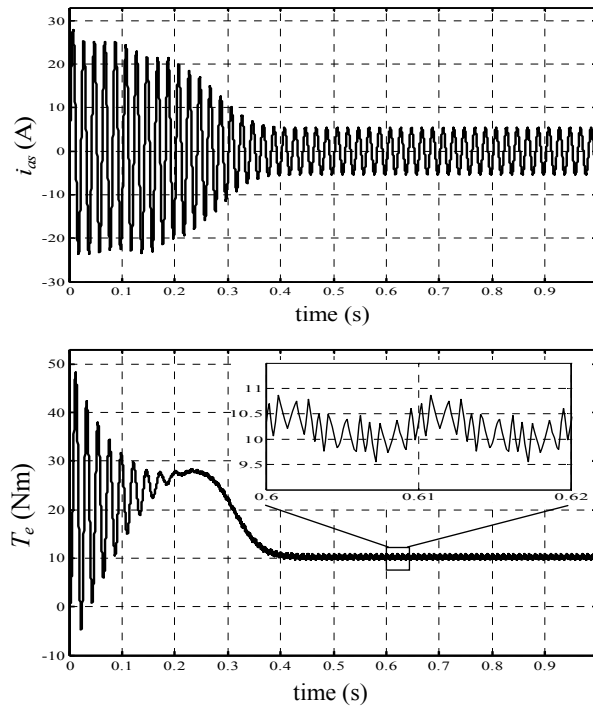


Figure 20

Stator current (top) and electromagnetic torque (bottom) of the asynchronous machine fed by a 9-level USAMI controlled by the proposed neural HES

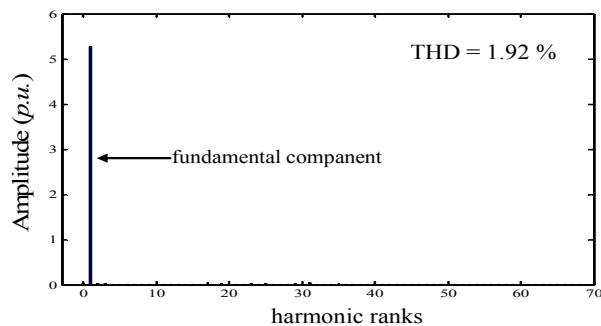


Figure 21

Frequency content of the stator current of the asynchronous machine fed by a 9-level USAMI controlled by the proposed neural HES

The results of the control based on the MCPWM are presented by Figs. 18 and 19. This first figure shows the stator current and the electromagnetic torque with significant fluctuations. The second figure shows the frequency content of the stator current. Results by using neural approach with the MLP issued for the

previous learning process are presented by Figs. 20 and 21. By comparing Fig. 19 to Fig. 21, it can be deduced that the neural HES efficiently cancels the harmonics of ranks 5, 7 and 11 from the output voltage V_{ab} . Moreover, the amplitudes of the harmonic distortions are very small compared to the amplitude of the fundamental component.

Performances obtained with both methods are summarized in Table 3. The THD measured on V_{ab} and resulting from the neural approach of the HES is smaller than the one obtained with the MCPWM method. The THD measured on the stator current i_{as} is reduced by a factor 2.63 with the neural HES compared to the MCPWM method. The control is thus optimized with the neural HES in order to avoid the asynchronous machine to absorb harmonics.

Table 3
Performances of the control methods

Control method	V_{ab} THD (%)	i_{as} THD (%)	f_{Cem} (Hz)	Δ_{Cem} (Nm)	Nb of θ_i
MCPWM	9.19	5.05	f	1.77	$2m = 48$
Neural HES	7.66	1.92	$2f$	1.31	$4p = 16$

It can also be seen that the electromagnetic torque continuously oscillates at a frequency f with the MCPWM method (because of the harmonics of rank 2 and 4 which are present in the output voltage). The torque oscillates at $2f$ with the neural approach. The neural method also reduced the number of switching angles by a factor 3 compared to the MCPWM method which is highly appreciated for the electronic devices.

Conclusions

The performance of motors fed by inverters are closely related to the strategy used to control its supplying inverter. Indeed, the control currents of the motor can be disturbed by harmonics introduced by the inverter and Harmonics Elimination Strategies (HES) are generally used. We propose a low-cost neural implementation of a HES to control a uniform step asymmetrical 9-level inverter. The approach is based on the learning and approximating of the relationship between the modulation rate and the switching angles with a Multi-Layer Perceptron. The resulting neural implementation of the HES uses very few computational costs. It is particularly well suited for real-time motor control tasks. The proposed neural approach is compared to the MCPWM strategy. Simulation results are given to show the high performance and technical advantages of the neural implementation of the HES for the control of a uniform step asymmetrical 9-level inverter. The proposed neural method efficiently cancels the current harmonic distortions in supplying an asynchronous machine. As a result, the torque undulations and the switching losses are significantly reduced.

Appendix

- Supply voltages of the partial inverters with:

pu Units: $u_{d1} = 1$, $u_{d2} = 1$ and $u_{d3} = 2$;

SI Units: $U_{d1} = 100\text{V}$, $U_{d2} = 100\text{V}$ and $U_{d3} = 200\text{V}$.

- Asynchronous machine data:

Stator resistance $R_s = 4.850\Omega$, Rotor resistance $R_r = 3.805\Omega$, Stator inductance $L_s = 0.274\text{H}$, Rotor inductance $L_r = 0.274\text{H}$, Mutual inductance $L_m = 0.258\text{H}$, Number of pole pairs $P = 2$, Rotor inertia $J = 0.031\text{kg.m}^2$, Viscous friction coefficient $K_f = 0.00136\text{ Nm.s.rad}^{-1}$.

References

- [1] Rodriguez, J., Lai, J. S., Peng, F.Z.: Multilevel Inverters: A Survey of Topologies, Controls, and Applications, in IEEE Transactions on Industrial Electronics, Vol. 49, No. 4, Aug. 2002, pp. 724-738
- [2] Manjrekar, M. D.: Topologies, Analysis, Controls and Generalization in H-Bridge Multilevel Power Conversion, Ph.D. thesis, University of Wisconsin, Madison, 1999
- [3] Mariethoz, S.: Etude formelle pour la synthèse de convertisseurs multiniveaux asymétriques: topologies, modulation et commande (in french), Ph.D. thesis, No. 3188, EPF-Lausanne, Switzerland, 2005
- [4] Ould Abdeslam, D., Wira, P., Merckl, J., Flieller, D., Chapuis, Y. A.: A Unified Artificial Neural Network Architecture for Active Power Filters, in IEEE Transactions on Industrial Electronics, Vol. 54, No. 1, Feb. 2007, pp. 61-76
- [5] McGrath, B. P., Holmes, D. G.: Multicarrier PWM Strategies for Multilevel Inverters, in IEEE Transactions on Industrial Electronics, Vol. 49, No. 4, Aug. 2002, pp. 858-867
- [6] Song-Manguelle, J., Mariethoz, S., Veenstra, M., Rufer, A.: A Generalized Design Principle of a Uniform Step Asymmetrical Multilevel Converter for High Power Conversion, in European Conference on Power Electronics and Applications, EPE'01, Graz, Austria, Aug. 2001, pp. 1-12
- [7] Taleb, R., Meroufel, A., Wira, P.: Commande par la stratégie d'élimination d'harmoniques d'un onduleur multiniveau asymétrique à structure cascade (in french), in Acta Electrotechnica, Vol. 49, No. 4, 2008, pp. 432-439
- [8] Chiasson, J. N., Tolbert, L. M., McKenzie, K. J., Du, Z.: A Unified Approach to Solving the Harmonic Elimination Equations in Multilevel Converters, in IEEE Transactions on Power Electronics, Vol. 19, No. 2, Mar. 2004, pp. 478-490

- [9] Haykin, S.: *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Upper Saddle River, N. J., 2nd edition, 1999
- [10] Taleb, R., Meroufel, A., Wira, P.: Harmonic Elimination Control of an Inverter Based on Artificial Neural Network Strategy, in 2nd IFAC International Conference on Intelligent Control Systems and Signal Processing (ICONS 2009), Istanbul, Turkey, Sept. 2009, on CD
- [11] Bishop, C. M.: *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995
- [12] Song-Manguelle, J. : *Convertisseurs multiniveaux asymétriques alimentés par transformateurs multi-secondaires basse-fréquence: réactions au réseau d'alimentation (in french)*, Ph.D. thesis, No. 3033, EPF-Lausanne, Switzerland, 2004
- [13] Dahidah, M. S. A., Agelidis, V. G.: Selective Harmonic Elimination PWM Control for Cascaded Multilevel Voltage Source Converters: A Generalized Formula, in *IEEE Transactions on Power Electronics*, Vol. 23, No. 4, Jul. 2008, pp. 1620-1630
- [14] Khomfoi, S., Tolbert, L. M.: Fault Diagnostic System for a Multilevel Inverter Using a Neural Network, in *IEEE Transactions on Power Electronics*, Vol. 22, No. 3, May 2007, pp. 1062-1069

Energetic Utilisation of Pyrolysis Gases in IC Engine

Viktória Barbara Kovács, Attila Meggyes

Department of Energy Engineering, Faculty of Mechanical Engineering, BME,
Műgyetem rkp. 3, H-1111 Budapest, Hungary
kovacsv@energia.bme.hu, meggyes@energia.bme.hu

Abstract: The use of alternative energy sources like pyrolysis gases as a source of renewable energy for combined heat and power generation could provide an effective and alternative way to fulfil remarkable part of the increasing energy demand of the human population as a possible solution of decentralized power generation. Therefore the role of utilization of pyrolysis gases rapidly grows in Europe and all around the world. The energetic utilization of these low heating value renewable gaseous fuels is not fully worked out yet because their combustion characteristics significantly differ from natural gas, and this way they are not usable or their utilization is limited in devices with conventional build-up. At the Department of Energy Engineering of BME the IC Engine utilization of pyrolysis gases was investigated. The power, efficiency, consumption and exhaust emission were measured and indication was made to determine the pressure and heat release in the cylinder at different engine parameters.

Keywords: renewable, pyrolysis gas, IC engine, power, efficiency, indication, heat release, exhaust emission

1 Introduction

This paper is focusing on the investigation of combustion characteristics of pyrolysis gases from the aspect of energetic utilization. The utilization of renewable alternative energy sources like liquid bio-fuels [1], [2] biogases and pyrolysis gases will have a major role in mitigating the climate change while the increasing energy demand of the humanity need to be fulfilled and the sustainable development should be maintained. Because renewable energy sources, among them bio- and pyrolysis gases used in CHP units could be an effective alternative to fulfil remarkable part of this energy demand as a possible solution of decentralized power generation because the total efficiency of a gas engine operated in cogeneration or trigeneration can be more than 90%. [3], [4]. Therefore the role of utilization of renewable gaseous fuels rapidly grows in Europe and all around the world. The renewal's share of the total energy sources

is below the expected in Hungary so not only the utilisation of biogases but the utilisation of other renewable gaseous fuels such as pyrolysis gases is recommended.

However several investigation was made to determine the combustion characteristic of these pyrolysis gases operating in various heat engines [5], [6], [7], but their energetic utilization is not fully worked out yet because of their different composition their combustion characteristics significantly differ from those conventional fuels like natural gas or PB gas [3], which are already used for power generation. So pyrolysis gases are not usable or their utilization is limited in heat engines with conventional build-up.

Therefore the energetic utilization of pyrolysis gases in IC engine with conventional build-up is problematic if their inert or hydrogen content is high. Therefore measurements were made to determinate the effect of the different composition, especially the high H₂ content of pyrolysis gases on the operation of IC Engine with conventional build-up at the Department of Energy Engineering – BME.

2 Properties of Pyrolysis Gases

The combustion characteristic of these renewable fuels differs from natural gas due to their different composition. The difference of pyrolysis gases and biogases it that biogases contain mainly CH₄ and CO₂ and an irrelevant amount of N₂, H₂, CO and SO₂, but pyrolysis gases beside CH₄ and CO₂ mainly contain CO and H₂ an depending on the production technology high amount of N₂. The LHV of these gases is low due to their high inert and /or high hydrogen content because the LHV per volume of hydrogen is much lower than the LHV of methane (Table 1). In case of gaseous fuel the LHV per volume is more important because the IC engine has constant mixture volume intake.

Table 1
LHV of hydrogen and natural gas

	Hydrogen	Natural gas
LHV [MJ/kg]	119.9	50.03
LHV [MJ/Nm ³]	10.78	35.9

For modelling the combustion quality of pyrolysis gases three different trial gas mixtures were determined, because the composition and the quality of pyrolysis gas mainly depends on its production technology. Pyrolysis gases can be gasified with outer heat source or with inner heat source, which could be air or pure oxygen. The H₂O content of these trial gas mixtures was neglected during the calculations and measurements, because it can be easily separated from the other components.

Table 2
Content and properties of different pyrolysis gases

component [V/V%]	“natural gas”	pyrolysis gases		
		outer heat source	inner heat source	
			air	oxygen
		“anaerobe pyrolysis gas”	“producer gas”	“synthesis gas”
CH ₄	100	8	5	3
H ₂	0	38	20	40
CO	0	20	20	40
CO ₂	0	20	5	17
N ₂	0	14	50	0
LHV [MJ/m ³]*	35.90	9.49	6.47	10.44
Wo [MJ/m ³]*	53.66	12.61	7.94	13.68

* calculated at 273 K and 101325 Pa

Apart from the LHV, the Wobbe number is a crucial parameter as far as combustion process of gaseous fuels is concerned, because it shows the changeability of gaseous fuels. By the changing of the gas composition the Wobbe number and accordingly the heat load of the combustion equipment changes too. The Wobbe number of these renewable gases significantly differs from the natural gas; therefore it is clear that the utilization of these renewable gases needs several investigations. [8]

From the point of view of stable operation of the engine the variation of these two parameters should be kept in the range of $\pm 5\%$ it is obvious that neither the LHV nor the Wobbe number can be kept in the required range in case of pyrolysis gas operation.

According to previous investigations gas type “producer gas” was chosen to do the measurements on IC engine, because the usual pyrolysis gas production technology is gasification with air as an inner heat source.

The combustion characteristic of “producer gas” differs from natural gas. The H₂ content of “producer gas is quite high, which is critical from the point of view of knock, and it has very low LHV which is critical from the point of view of power. Due to the high H₂ content of “producer gas” the direct use in IC engine was not recommended therefore the “producer gas” was mixed to natural gas. The combustion properties of these fuel mixtures was investigated. In case of the calculations the natural gas was modelled by pure methane gas, because the natural gas type “D” that is provided in Hungary contains more than 98 V/V% methane.

Theoretical calculation were made with CHEMKIN 4.0 software - GRI 3.0 mechanism, which is a reduced mechanism for modelling methane combustion and is capable for modelling the combustion of CO, H₂ and the formation of NOx

too. Therefore it is capable for modelling the combustion of pyrolysis gases. The two most important combustion parameters were calculated: the adiabatic flame temperature (T_{ad} [K]) and the laminar flame velocity (u [cm/s]). The effect of the admixed “producer gas” was investigated on these combustion parameters.

The “producer gas” has lower adiabatic flame temperature than natural gas, but the laminar flame velocities of “producer gas” and natural gas are quite the same in the operation range of a gas engine. Therefore by the increasing “producer gas” content of the fuel mixture the adiabatic flame temperature decreases, but the change of the laminar flame velocity is not relevant, because it do not exceeds $\pm 2\%$ of the laminar flame velocity of natural gas. But it slightly decreases until 40% “producer gas” content and above it increases until 90% “producer gas” content (Figure 1). [11]

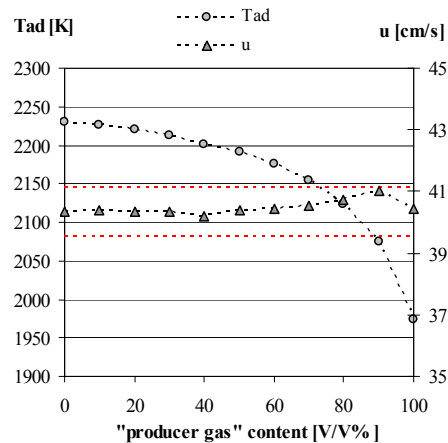


Figure 1

Calculated adiabatic flame temperature and laminar flame velocity against “producer gas” content at $\lambda=1$, 273 K, 10135 Pa

Two dimensionless factors were defined to determine the effect of the “producer gas” content of fuel mixture on the combustion properties [9]. The LHV ratio (γ) shows how many times higher the LHV of natural gas is compared to “producer gas – natural gas mixtures.

The next dimensionless factor, the theoretical fuel-air mixture volume ratio (ϵ) shows how many more times biogas can be used compared to natural gas to keep the excess air ratio of 1 m³ fuel - air mixture constant.

The value of these two factors (γ , ϵ) depends on “producer gas” content of the fuel mixture. If the LHV ratio (γ) and the theoretical fuel-air mixture volume ratio (ϵ) are equal at a given “producer gas” content, the LHV decrement caused by the low LHV of “producer gas” can be equalized by the increasing fuel proportion of the fuel-air mixture.

Figure 2 shows that the values of γ and ε are nearly the same until 50 V/V% “producer gas” content, but above it the decrement of LHV is higher than the possible increment of the air-fuel mixture volume flow so the effect of the decreasing heating value could not be equalized. This phenomenon was confirmed by the following measurements.

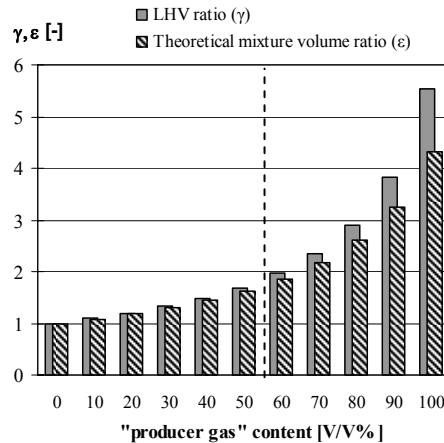


Figure 2

LHV- and theoretical fuel-air mixture volume ratios of different “producer gas” – natural gas mixtures calculated at 273 K and 10135 Pa

3 Experimental Set-up

Measurements were made at the laboratory of the Department of Energy Engineering of BME to determine the combustion characteristic of biogases on a BAG-30 gas engine unit which was modified for laboratory measurements (Figure 3).

The “producer gas” was modelled by a trial gas mixture that was ordered from Linde Gas Hungary in a bundle and was mixed through a multistage pressure regulator to the natural gas and the mixture was aspirated by the engine. The homogenization of the mixture was prepared in a mixing unit. The composition of the mixture was controlled with a CH_4 analyzer.

The control of the gas engine was made with the asynchronous generator of the engine. The constant speed was provided by a frequency inverter which was connected to the asynchronous generator. The electric power was measured with the frequency inverter. During the measurements intake pressure was kept at a constant value.

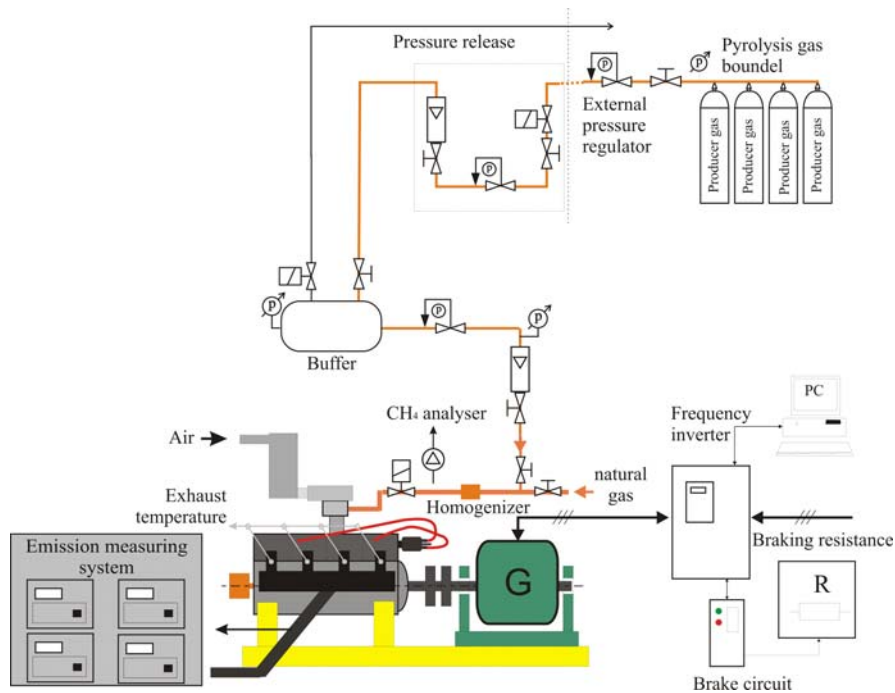


Figure 3
Build-up of the measuring system

The evolved pressure was measured with a piezo pressure transducer in a Kistler 6517-A spark plug which was installed in the 1st cylinder of the engine. Test series consist of 100 combustion cycles with sampling rate of 1024 per cycle and were averaged by statistical methods.

The gross heat release in the cylinder was calculated from the combustion pressure with a software which was developed at the Department of Energy Engineering [10].

The emissions of the gas engine were measured with a Horiba MEXA-8120F emission measuring system. The oxygen content of the exhaust gas which was needed for the determination of the excess air ratio (λ) was measured by a SERVOMEX 570A oxygen analyzer. The measured data of the engine was recorded by electronic data collection system.

The reference measurements were made with natural gas (0 V/V% “producer gas” content). The measurements were made at 10; 20; 30; 40; 50 and 60 V/V% “producer gas” contents. At higher “producer gas” content the operation of engine become unstable, so with higher than 60 V/V% “producer gas” content measurement could not be made. The impact of the increasing “producer gas” content was investigated on the engine parameters: power, efficiency,

consumption, and emission. Due to the comparability and reproducibility the measurements were made at constant spark timing, speed and intake pressure in case of several excess air ratios.

4 Results

From the point of view of engine operation the in-cylinder peak pressure is very important parameter (Figure 4), because it affects the power of the engine.

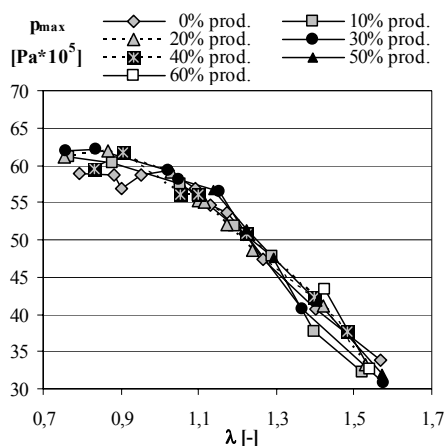


Figure 4

Measured maximum pressures in the cylinder against excess air ratio

It is observable that the peak pressures are quite the same by increasing “producer gas content of the fuel mixture, but the operation range of the engine narrows as well. Namely in case of 50 V/V% “producer gas the operation range is only the half of the operation range of the reference measurement and it is shifted to leaner mixtures. In case of 60 V/V% “producer gas under $\lambda=1.4$ measurements could not be made, because the operation of the engine was unstable.

In order to compare the form of the cylinder pressures Figure 5 shows the normalized measured pressures in the cylinder at constant excess air ratio ($\lambda=1.4$). It is observed that neither relevant change of the peak pressures nor relevant shift of the peak pressure from the TDC (which was adjusted to 360 degree) could be experienced.

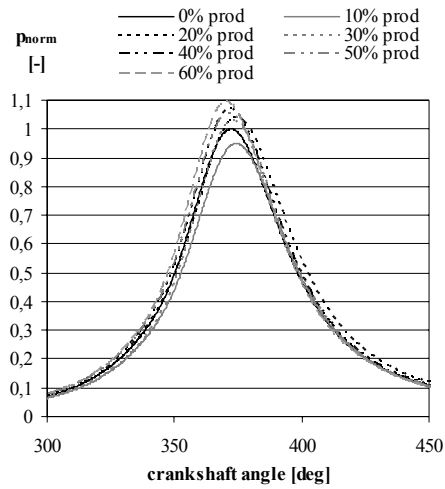


Figure 5

Cylinder pressures at $\lambda=1,4$ in case of different “producer gas” content.

Figure 6 shows the normalized calculated heat release rate in the cylinder. It is observed that in case of all “producer gas” content the maximum heat release rate is up to 10% higher than the maximum heat release rate of natural gas.

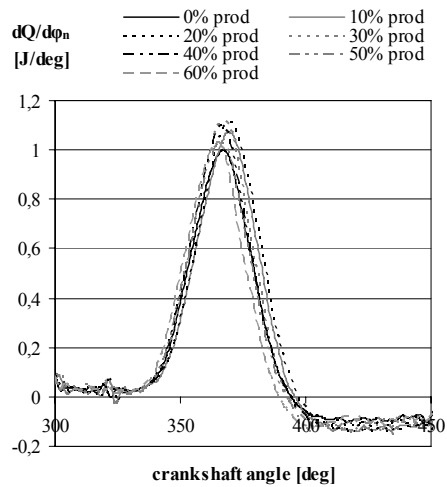


Figure 6

Calculated heat release gradient at $\lambda=1,4$.

To better visualise the impact of “producer gas” on heat release Figure 7 shows the normalised maximum heat release rate and Figure 8 shows the inherent crankshaft angles against the “producer gas” content of the fuel mixture in case of different excess air ratios. The base of the normalisation was the maximum heat release rate of the reference gas at stoichiometric mixture.

Figure 7 shows, that the maximum heat release rate decrease with the increase of the oxidiser (air). The decrement is non linear. Despite the decrease of the adiabatic flame temperature the increase of “producer gas” content does not involve the decrement of the maximum heat release rate at constant excess air ratio; moreover it is well observable, that in case of all “producer gas” contents the maximum heat release rate is the same or higher than the maximum heat release rate of the reference gas. However at constant excess air ratio relevant change or tendency could not be observed.

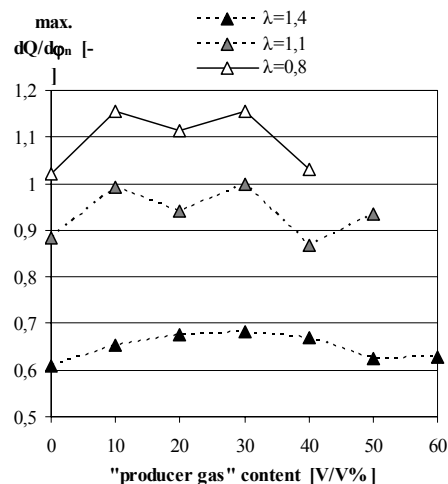


Figure 7

Maximum heat release rate against “producer gas” content at different excess air ratios

Figure 8 shows, that the location of maximum heat release rate shifts further from the TDC with the increase of the oxidiser (air). The shift is non linear. The curves are in correlation with the calculated laminar flame velocity (Figure 1).

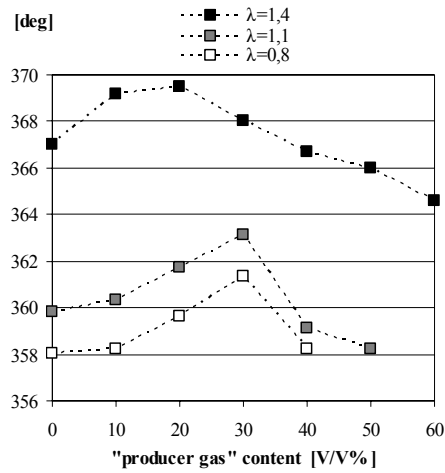


Figure 8

The location of the maximum heat release rate against "producer gas" content at different excess air ratios

As it was determined at the theoretical calculation in case of stoichiometric mixtures the change of the laminar flame velocity is not significant but shows moderate decrease until 30-40 V/V% "producer gas" content and increase above it. That is in good correspondence with the location of the maximum heat release rate, because as the laminar flame velocity decrease the location of the maximum shifts further from the TDC.

Although the LHV of "producer gas is very low due to its low L_0 the heat input could be kept during the whole measuring range in case of all "producer gas" contents (Figure 9). Necessarily the consumption of the mixture needed to be increased (Figure 10).

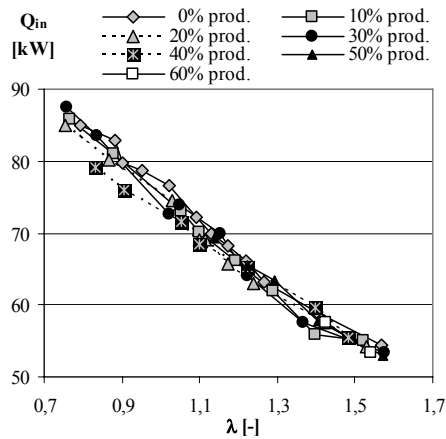


Figure 9

Heat input of different „producer gas” natural gas mixtures

It is observed that not only the consumption increases but the operation range of the engine narrows and shifts to leaner mixtures the by the increasing “producer gas” content of the mixture (Figure 10). Namely in case of 50 V/V% “producer gas” the operation range is only the two third of the operation range of the reference gas.

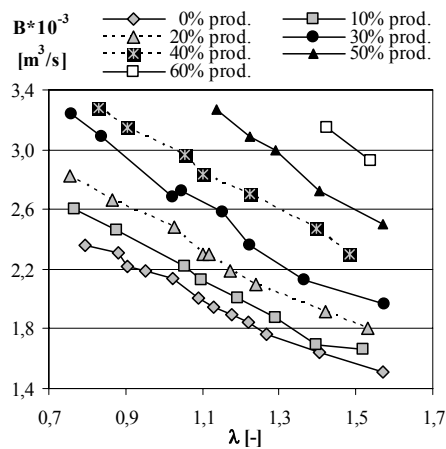


Figure 10

Consumption of different „producer gas” natural gas mixtures

Due to the increasing consumption relevant effective power change could not be experienced in the whole measuring range (Figure 11).

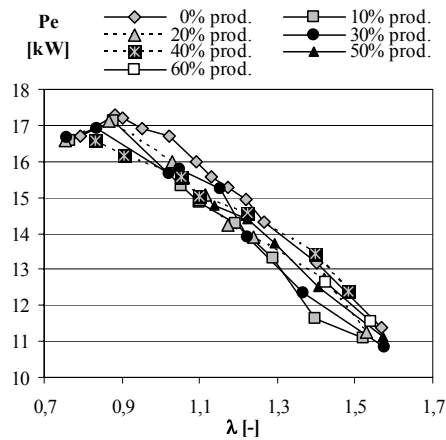


Figure 11

Effective power of different „producer gas” natural gas mixtures

The change of the effective efficiency is not relevant either, but above $\lambda=1$ the deviation of the values is remarkable (Figure 12).

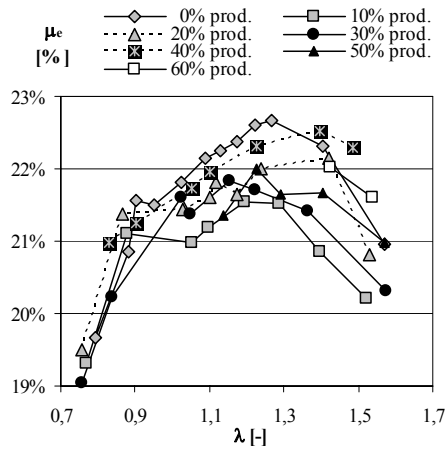


Figure 12

Effective efficiency of different „producer gas” natural gas mixtures

The results of the exhaust gas emission measurements turned out as expected [12].

Relevant change in the CO₂ emissions could not be observed (Figure 13). The shapes of the curves of CO₂ emissions are in good correspondence with the curves of power. The maximum CO₂ emissions are around $\lambda \approx 0.9$ where the power maximums lay. In case of richer and leaner mixtures the CO₂ emission decreases.

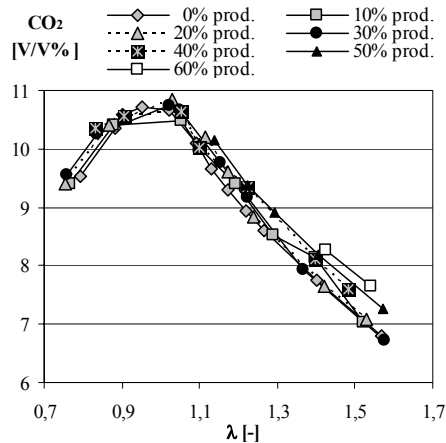


Figure 13

Measured CO₂ emission of different „producer gas” natural gas mixtures

In case of the determination of NO_x (NO, N₂O, NO₂) emission only the NO emission was measured, because NO_x contains more than 95% NO. Alike in case of the CO₂ emissions relevant change in the NO emissions could not be observed (Figure 14). The shape of the curves of NO emission is acceptable. The NO maximum are around $\lambda = 1.1$ and decreases both in case of lower and higher excess air ratios.

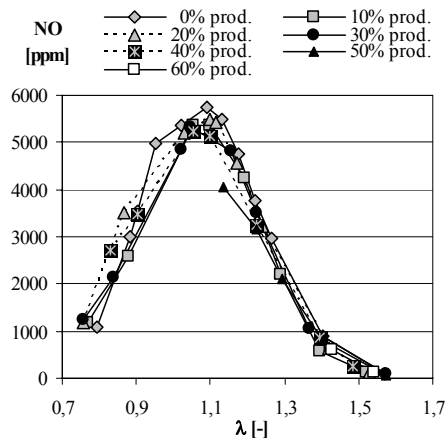


Figure 14

Measured NO emission of different „producer gas” natural gas mixtures

The change in the THC emissions is not relevant. The shape of the curves of the THC emissions is acceptable (Figure 15).

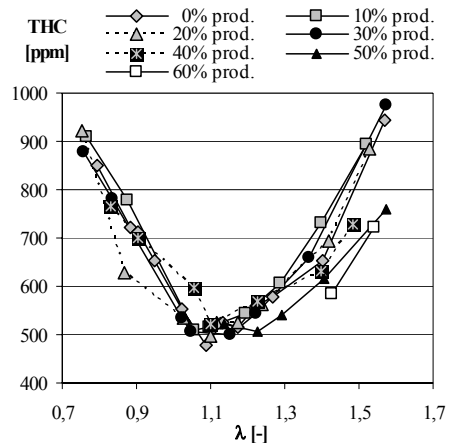


Figure 15

Measured THC emission of different „producer gas” natural gas mixtures

The minimum of THC emissions is around $\lambda=1.1$ and they increase both in case of lower and higher excess air ratios as incomplete combustion takes place. That is in good correspondence with the NO emissions. In case of lean mixtures the THC emission slightly decrease with the increasing “producer gas” content of the fuel mixture, because due to the higher hydrogen content the flame propagation velocity increases so the hydrocarbon content of the fuel can be combusted to CO, but the combustion could not be completed due to the freezing chemical reactions.

Alike in the case of THC emission the change of CO emissions is not relevant either (Figure 16). The shape of the curves of CO emissions formed also as they were expected. In case of lean mixtures the CO emission slightly increases with the increasing “producer gas” due to the freezing chemical reactions; and in case of enriching the fuel – air mixtures the it increases considerably due to the absence of oxidizer (air) caused incomplete combustion.

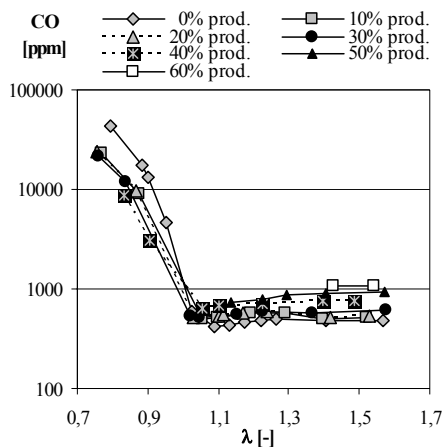


Figure 16

Measured CO emission of different „producer gas” natural gas mixtures

Conclusions

According to previous investigations above 50% “producer gas” content power decrement and knock was expected [12].

However neither relevant change of the investigated parameters nor knock could be experienced, but due to the low LHV of “producer gas” the consumption increased.

But from the measurements of “producer gas” - natural gas mixtures can be set out that the energetic utilization of “producer gas” in IC engine with conventional build-up is limited because of the low LHV of these gaseous fuels, above a given “producer gas” content the combustion could not take place. Accordingly above 40 V/V% “producer gas” content the operation range of the engine narrows and shifts to leaner mixtures by the increasing “producer gas” content of the fuel mixture; and at the given operation conditions above 60 V/V% “producer gas” content the engine was unable to run on fuel with such high “producer gas” amount.

Therefore above 40 V/V% producer gas content and especially in case of pure pyrolysis gas operation the IC engine need to be adjusted to the used pyrolysis gas to avoid considerable losses, e.g. spark timing or the mixing unit need to be modified.

These results are in good correspondence with the others in the referred literature; however the compositions of the investigated pyrolysis gases are different in case of each researches.

References

- [1] Laza T., Bereczky Á.: Determination of the Evaporation Constant in Case of Pure and with Alcohol Mixed Rape Seed Oil., in Proceedings of 16th International Conference in Mechanical Engineering, Brassov, May 1-4, 2008, pp. 232-237
- [2] Bereczky Á.: Utilisation of Bio Fuels in Internal Combustion Engines, in Proceedings of 8th Conference on Heat Engines and Environmental Protection, Balatonfüred, May 28-30, 2007, pp. 43-47
- [3] Meggyes A., Bereczky Á.: Energetic Analysis of Combined Gas Engine Systems, *Energetika*, 2007/3, pp. 18-22 (in Hungarian)
- [4] Nagy V., Meggyes A.: Utilization of Biogas in Gas Engines, in Proceeding of 8th International Conference on Heat Engines and Environmental Protection, Balatonfüred, 2007, pp. 95-100
- [5] Andoa Y., Yoshikawaa K., Becka M., Endo H.: Research and Development of a Low-BTU Gas-driven Engine for Waste Gasification and Power Generation, *Energy*, Vol. 30, pp. 2206-2218, 2005
- [6] Ramadhas S. A., Jayaraj S., Muraleedharan C.: Power Generation Using Coir-Pith and Wood Derived Producer Gas in Diesel Engines, *Fuel Processing Technology*, Volume 87, Issue 10, pp. 849-853
- [7] Sridhar G., Sridhar H. V., Dasappa S., Paul P. J., Rajan N. K. S., Mukunda H. S.: Development of Producer Gas Engines, in Proceedings of the Institution of Mechanical Engineers Part D – Journal of Automobile Engineering, Volume 219, Issue D3, pp. 423-438
- [8] Kovács V. B., Meggyes A.: Effect of Different Gas Compositions and Combustion Circumstances on the Operation of Heat Engines, in Proceeding of 7th International Conference on Heat Engines and Environmental Protection, Balatonfüred, May 23-25, 2005, pp. 88-94
- [9] Kovács V. B., Meggyes A., Bereczky Á.: Investigation of Utilization of Pyrolysis Gases in IC engine, in Proceeding of Sixth Conference on Mechanical Engineering, Budapest, May 29-30, 2008, CD
- [10] Lukács K.: Development of a Dual Fuels Diesel Engine System for Power Generation, Diploma, Dep. Energy Eng., BME 2007 (in Hungarian)
- [11] Kovács V. B., Meggyes A.: Investigation of Utilisation of Pyrolysis Gases in IC Engine, in Proceedings of X. Energetika-Elektrotechnika Konferencia – ENELKO, Marosvásárhely, Románia, October 8-11, 2009, pp. 93-98 (in Hungarian)
- [12] Kovács V. B., Meggyes A.: Investigation of IC Engine Utilization of Renewable Gases, Proceedings of 24. Deutscher Flammentag, Bochum, September 16-17, 2009, pp. 519-522