NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS

ABSTRACTS OF DOCTORAL DISSERTATIONS

Athens 2012

Volume 7

NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS

ABSTRACTS OF DOCTORAL DISSERTATIONS

ISSN: 1791-7948

Volume 7

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
Panepistimiopolis, 15784 Athens, Greece

PREFACE

This volume contains the extended abstracts of the Doctoral Dissertations conducted in the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens and completed in the time period 12/2010 to 12/2011.

The goal of this volume is to demonstrate the breadth and quality of the original research conducted by our Ph.D. students and to facilitate the dissemination of their research results. We are happy to present the 7th collection of this kind and expect this initiative to continue in the years to come. The submission of an extended abstract in English is required by all graduating Ph.D. students.

We would like to thank the Ph.D. students who contributed to this volume and hope that this has been a positive experience for them. Finally, we would like to thank our graduate student Mr. Nikos Bogdos for his help in putting together this publication. The painting shown in the cover is "*Ταξείδι...*" by artist *Διονύσης Καμπόλης*.

The Committee of Research and Development

A. Eleftheriadis
M. Koubarakis
E. Manolakos (Chair)
T. Theoharis

Athens, July 2012

Table of Contents

Text Detection in Images and Videos

Marios Anthimopoulos [*]

[1]Department of Informatics and Telecommunications
National and Kapodistrian University of Athens

[2]Computational Intelligence Laboratory
Institute of Informatics and Telecommunications
National Centre for Scientific Research "Demokritos

anthimop@iit.demokritos.gr

Abstract. The goal of a multimedia text extraction and recognition system is filling the gap between the already existing and mature technology of Optical Character Recognition and the new needs for textual information retrieval created by the spread of digital multimedia. A text extraction system from multimedia usually consists of the following four stages: spatial text detection, temporal text detection – tracking (for videos), image binarization – segmentation, character recognition. In the framework of this PhD thesis we dealt with all the stages of a multimedia text extraction system, focusing though on the designing and development of techniques for the spatial detection of text in images and videos as well as methods for evaluating the corresponding result. Two methods for the evaluation of the text detection result were proposed that deal successfully with the problems of the related literature. Each of them uses different criteria while both of them are based on intuitively correct observations. Finally, a very efficient method was developed for the temporal detection of text which actually conduces to a better spatial detection while concurrently enhances the quality of the text image.

Keywords: text detection, text recognition, artificial text, scene text, natural scene images, Video OCR, multimedia information retrieval, evaluation.[1]

1 Introduction

Nowadays the size of the available digital video content is increasing rapidly. This fact leads to an urgent need for fast and effective algorithms for information retrieval from multimedia content. Textual information in multimedia constitutes a very rich source of high-level semantics for retrieval and indexing. Document image processing, after many decades of research, has reached a high level of text recognition accuracy, for traditional scanner-based images. However, these techniques fail to deal with text appearing in videos or camera-based images. The goal of a multimedia text extraction and recognition system is filling this gap between the already existing and mature technology of Optical Character Recognition and the new kinds of data.

Mainly, there exist two kinds of text occurrences in videos and images, namely artificial and scene text. Artificial text, as the name implies, is artificially added in order to describe the multimedia content while scene text is textual content that was captured by a camera as part of a scene, e.g. text on T-shirts or road signs. Figure 1(a) presents a video frame with artificial text and Figure 1(b) a natural scene image containing scene text. Text can also be classified into normal or inverse. Normal is denoted any text whose characters have lower intensity values than the background while inverse text is the opposite. In Figure 1(b), "FIRE" is considered as normal while the rest of the text is considered inverse.

[1] [*]Dissertation Advisors: [1]Sergios Theodoridis, Professor – [2]Dr. Basilis Gatos, Researcher

(a) (b)

Figure 1. (a) Video frame with artificial text, (b) Image with scene text

2 Related Work

Several methods for text extraction from multimedia have been proposed the last decade. Most of them are divided in the following stages: spatial text detection, temporal text detection, image binarization – segmentation and character recognition. From these stages the most crucial is the stage of spatial detection which actually concentrates the focus of most researchers in this field. The performance of the whole text extraction system depends on the accurate localization of text in an image or video frame. Much lesser work has been done for the temporal detection of text in videos while usually only static text is considered. Some methods have also been proposed for the binarization of the text image, although most researchers use state-of-the-art algorithms from the classic document analysis research area. Finally a very important aspect that has not been sufficiently studied is the development of the corresponding evaluation protocols which are necessary for optimizing the algorithms as well as comparing the several methods in literature. In this section we will outline the techniques found in literature for the different stages of text extraction from images and video frames.

In general, the existing spatial text detection methods can be roughly divided in two categories: region-based and texture-based. Region-based methods group pixels that belong to the same character based on the colour homogeneity, the strong edges between character and background or by using a stroke filter. Then, the detected characters are grouped to form textlines according to colour, size and geometrical rules. Texture-based algorithms scan the image at different scales using a sliding window and classify image areas as text or non-text based on texture-like features. Another possible categorization of text detection methods could be dividing them into heuristic and machine learning techniques. Typical heuristic, region-based approaches that rely on connected components can be found in [1-3]. Other region-based methods detect text based on edge or stroke information, i.e. strength, density or distribution [4-7]. DCT coefficients globally map the periodicity of intensity images and they have been widely used as texture features for heuristic texture-based methods [8-12]. Some hybrid methods have also been proposed [13-18]. Hybrid techniques combine the efficiency of a heuristic coarse stage with an accurate machine-learning refinement stage. Several pure machine learning, texture-based approaches have also been proposed for the detection of text areas with great success. These methods use directly machine learning classifiers to detect text [19-23]. The main shortcoming of the methods attributed to this category is the high computational complexity since a sliding window is required to scan the entire image, requiring thousands of calls to the classifier per image.

Very few researchers have focused n the problem of the temporal text detection mainly due to the objective difficulties of the task, the lack of an unbiased evaluation methodology and the great effort that is needed for the ground truth annotation. These existing attempts consider mainly artificial text and use standard object tracking methods while the result is usually evaluated optically. Lienhart et al. [24] considered text static or linearly moving and applied tracking with block matching using the least mean square criterion. Antani et al. [25] and Gargi et al. [11] use motion vectors in compressed MPEG-1 videos based on the work of Nakajama et al. [26] and Pilu et al. [27]. Li et al. [28] use affine transformation and Sum of Square Difference – SSD for the detection of moving text followed by a validation process which is based on Mean Square Error – MSE) and the moving text trajectory.

Within the framework of multimedia text extraction some binarization methods have also been proposed. Text detection methods based on connected components often performed image binarization

followed by heuristic grouping of character components [2,29]. LeBourgeois et al. [30] applied the Maximun Entropy Method combined with projection analysis. Antani et al. [31] use the binarization method proposed by Kamel and Zhao [32]. Lienhart and Wernike use a global binarization method based on text and background color estimation for each bouding box [22]. Wu, Manmatha et al. [33] calculate the threshold from text chips while Chen et al. [34] propose asymmetric Gabor filters for character stroke enhancement. Chen et al. [35] propose also a Markov Random Field (MRF) based on 2x2 cliques. Finally, many researchers use classic binarization algorithms like Otsu [36], Ohya [37], Niblack [38] and Sauvola [39].

The evaluation of a text detection system is an aspect not as trivial as it might seem. It resembles the generic problem of object detection evaluation having additionally its own issues. Most researchers use for their experimentation simple boxed-based or area-based methods while very few works have focused on the specific problem of evaluation. In [40] and [41] Kasturi, Manohar et al. propose as overall measure of text detection in a frame, a box-based measure called Frame Detection Accuracy (FDA). The evaluation methods of this kind are based on the mapping between ground truth and detected objects. Especially for the text detection problem, text lines are considered to be the objects where a text line is usually defined as an aligned series of characters with a small intermediate distance relative to their height. However, this subjectively small distance can result arbitrarily in bounding box splits or merges among annotators and detectors making the object mapping inappropriate. In addition, the number of correctly retrieved boxes is not generally a measure of the retrieved textual information since the number of characters in different boxes may vary considerably. Wolf et al. [42] proposed the creation of match score matrices with the overlap between every possible pair of blocks, in order to evaluate document structure extraction algorithms. The benefit of this kind of algorithms is their ability to consider the possible splits or merges of the bounding boxes besides one-to-one matching. However, in order to match two ground truth boxes with one resulting box, the total overlap threshold has to be very low (~40%). This will have as a result accepting as correct, a box with size even higher than the double size of the ground truth box. Many researchers have used the overall overlap to compute area-based recall and precision measures ([5], [13]). However the main drawback here similarly to the box-based approaches is the fact that the number of retrieved pixels does not correspond to proportional textual information since different textlines may contain characters of various sizes.

3 Spatial Text Detection

In the framework of this thesis we proposed three algorithms for text spatial detection - localization in images or/and video frames. Firstly, we proposed an edge-based heuristic method for the localization of artificial text in video frames. The second text detection system we developed is a two-stage scheme that uses the previous method as a first stage while a second machine, learning texture-based stage follows refining the initial result. The third detector we proposed uses directly machine learning techniques to distinguish text from background using texture-like features and is capable of detecting both artificial and scene text in images and video frames.

3.1 Edge-based heuristic method

The proposed algorithm [43] (Fig. 2) exploits the fact that text lines produce strong vertical edges horizontally aligned and follow specific shape restrictions. Using edges as the prominent feature of our system gives us the opportunity to detect characters with different fonts and colors since every character present strong edges, despite its font or color, in order to be readable. First, an edge map is created using the Canny edge detector [44]. Then, morphological dilation and opening are used in order to connect the vertical edges and eliminate false alarms. Bounding boxes are determined for every non-zero valued connected component, consisting the initial candidate text areas. Finally, an edge projection analysis is applied, refining the result and splitting text areas in text lines. The whole algorithm is applied in different resolutions to ensure text detection with size variability. Experimental results prove that the method is highly effective and efficient for artificial text detection in video frames with relative smooth background.

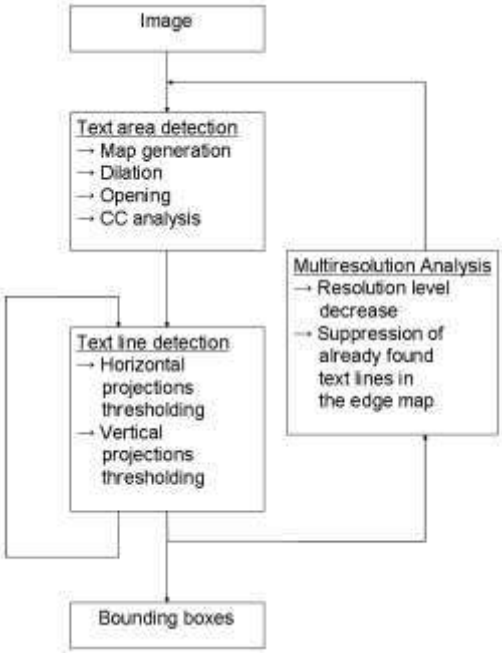


Figure 2. Flowchart of the proposed edge-based heuristic method for artificial text detection

3.2 Two-stage text detection scheme

This method consists of an initial coarse stage and a machine learning refinement stage [45]. In the first stage, text lines are detected based on the algorithm presented in section 3.1, leading in a high recall rate with low computational time expenses. In the second stage (Fig. 3), the result is refined using a sliding window and an SVM classifier trained on features obtained by a new Local Binary Pattern-based operator (eLBP) that describes the local edge distribution.

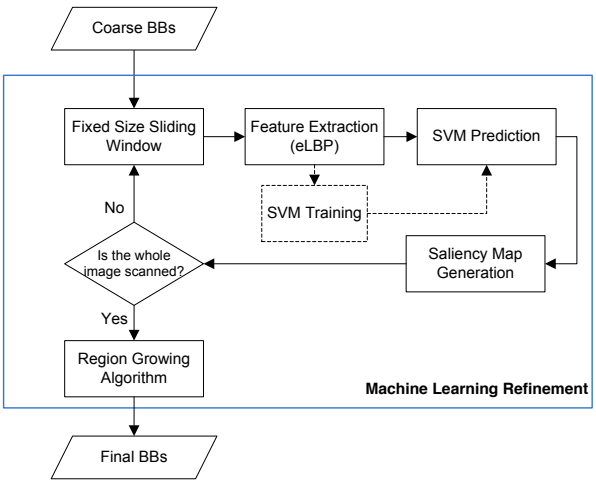


Figure 3. Flowchart of the proposed machine learning refinement stage

The feature set consists of the histogram values of the eLBP map which corresponds to each image area. In eLBP, a neighbouring pixel is represented by 0 if it is close to the center pixel or 1 if not. In order to define closeness, we require a minimum absolute distance *e* from the center to give to the pixel the binary value 1 (Figure 10).
Formally, the new eLBP operator is defined as:

$$\mathrm{eLBP}(x_c, y_c) = \sum_{n=0}^{7} S_e(i_n - i_c)2^n \qquad (1)$$

where function $S_e(x)$ is defined by:

$$S_e(x) = \begin{cases} 1\,, & |x| \geq e \\ 0\,, & |x| < e \end{cases} \tag{2}$$

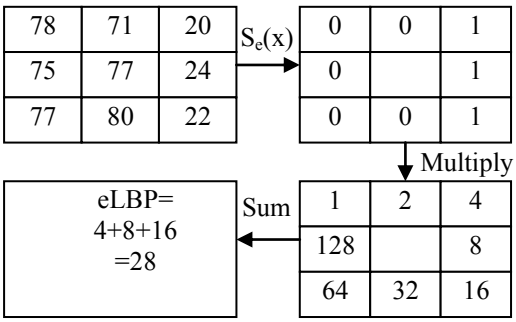


Figure 4. Example of eLBP computation

The value of *e* has to be large enough in order to avoid the arbitrary intensity variations caused by noise and small enough to detect all the deterministic intensity changes of texture. In [46] a value near 20 was proved to be satisfactory, after relative experimentation. Although this value was optimal for the discrimination of most text and non-text patterns there were still problems. Some text patterns of low contrast images presented edges that did not exceed the specified threshold and thus, were classified as non-text while non-text patterns of high contrast images presented strong enough edges to be classified as text. In order to solve this problem we propose the generation of multilevel eLBP edge histograms with different values for *e,* which will describe the edge distribution in different detail levels.These values for *e* are given by the quantile function of the exponential PDF which is denoted as:

$$e(i) = -M * \ln(1 - i/(L+1)) \tag{3}$$

where M is the mean value of the gradient image, i=1…L and L the number of different levels. For our experiments we used 8 number of levels (L=8) which resulted in 2048 features and forced us to do a feature reduction.

3.3 Machine learning method for artificial and scene text detection

In this work [47], we propose the use of a Random Forest within a sliding window model for the discrimination of text areas in the first stage, and then apply a gradient-based algorithm to achieve separation and refined localization of the text lines (Figure 2). This is actually a hybrid scheme combining an initial machine learning, texture-based technique with a heuristic, region-based refinement. The algorithm described above constitutes a fixed-scale detector so it is applied in multiple resolutions in order to detect text of any size. After processing all the resolutions needed, the final text line bounding boxes have been computed. Then, text lines are binarized and optionally segmented into words based on the distances between the resulting connected components. This segmentation is applied in cases where the text detection targets words instead of text lines.

The main contributions of this work is the choice of the classifier which provides efficiency and generalization capabilities, together with an improved, highly discriminative feature set that was designed particularly for reflecting the textual characteristics. Moreover, the use of a machine learning architecture for the first and most crucial stage, produces a generic and robust system for the detection of artificial and scene text in camera-based images and video frames. The description of the edge spatial distribution is done with the new MACeLBP (Multilevel Adaptive edge Local Binary Patterns) operator which considers the valuable RGB color information while it adapts the contrast levels locally to each area of the image producing an actual parameter-free feature set. The use of the SVM in [45] forced us to invent a reduced version of the feature set in order to have an efficient system. This fact resulted in loss of information for the description of the textual texture. Contrarily, in this work the use of a Random Forest and its capability to deal efficiently with high dimensional feature spaces allowed us to use the whole proposed feature set instead of a reduced version so all available information is exploited for a better description of texture. Finally, the impressive efficiency of the Random Forest gave us the capability to use it as the first and basic stage for scanning the whole image and detecting

text instead of just refining the coarse results of a heuristic and unreliable stage.

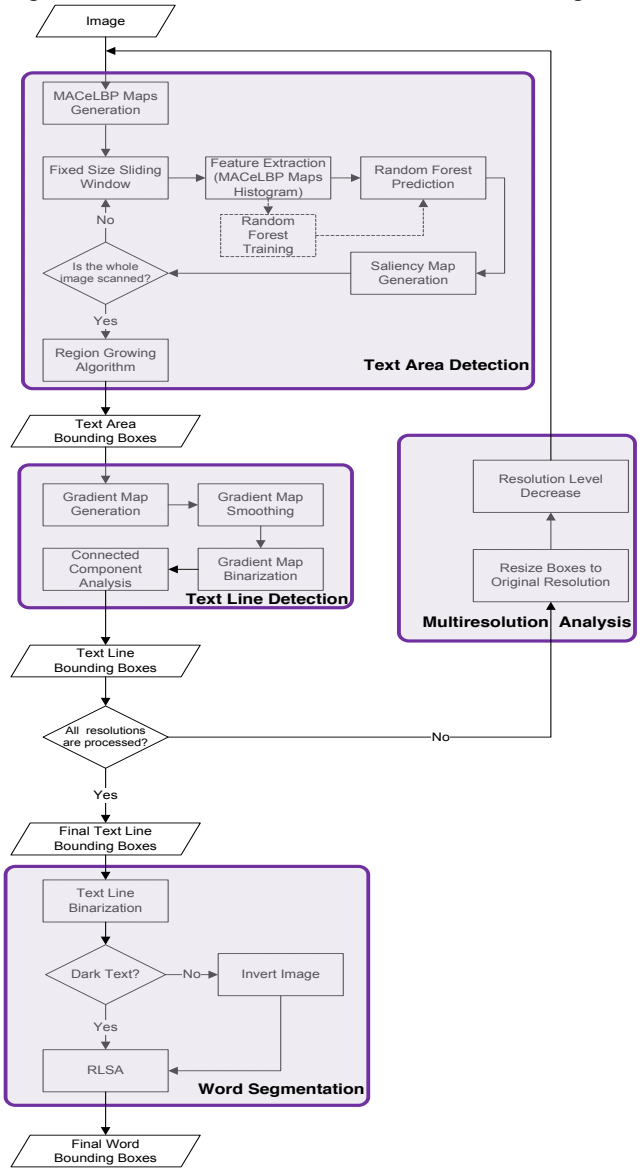


Figure 5. Flowchart of the proposed machine learning text detection algorithm

4 Temporal Text Detection

In order to obtain textual information from video, text detection and recognition in a single frame basis is not adequate. There is a giant amount of temporal information which has to be exploited. Every text line has to remain in the video for at least 2 seconds to be readable, which means at least 50 frames. Using this extra information we can remove the noise from the image and also smooth the background. Artificial text in television captured videos is usually static. Making this assumption we have the opportunity to use temporal information not only for enhancing the already detected text image but to improve text detection as well. Video frames are firstly averaged temporally and then fed to the text detection algorithm. Figure 6 presents an example of temporal text detection.

The algorithm mainly consists of three steps:

1. Temporal averaging of video (20 frames depth).
2. Text detection in the averaged video every 5 frames.
3. Matching of the detected text from frame to frame if overlap between the two boxes is over 80%.

Figure 6. Example of temporal text detection

5 Text Image Binarization

The proposed binarization method consists of an Otsu thresholding [36] followed by a normal/inverse text detection. In order to classify between normal or inverse text we firstly apply a connected component analysis. The numbers of white (*WCC*) and black (*BCC*) connected components are counted, discarding components with height less than 8 pixels or less than the 40% of the box height. If |*WCC*-*BCC*|>1 then the color that corresponds to the largest number of connected components is regarded as text color. Else if the distance between *WCC* and *BCC* is less or equal to 1, the condition for the inversion is based on the pixel values of the borders of the bounding boxes. If the majority of border pixels are black then text is considered inverse. Finally, color inversion is applied for every text detected to be inverse so the final result will be normal binary text.

6 Text Detection Evaluation

6.1 Evaluation based on estimated number of characters

Ideally a text detection method as a part of a text extraction system should not be evaluated on the size of detected areas nor the number of detected boxes but on the number of the detected characters. Unfortunately, the number of characters in a bounding box cannot be defined by the algorithm but it can be approximated by the ratio width/height of the box, if we assume that this ratio is invariable for every character, the spaces between different words in a text line are proportional to its height and each textline contains characters of the same size.

In that way, the evaluation will be based on the recall and precision of the area coverage, normalised by the approximation of the number of characters for every box [45]. The overall metric will be the weighted harmonic mean of precision and recall also referred as the F-measure.

$$\text{Recall}_{ecn} = \frac{\sum_{i=1}^{N} \frac{|\text{GDI}_i|}{\text{hg}_i^2}}{\sum_{i=1}^{N} \frac{|\text{GB}_i|}{\text{hg}_i^2}} \qquad \text{Precision}_{ecn} = \frac{\sum_{i=1}^{M} \frac{|\text{DGI}_i|}{\text{hd}_i^2}}{\sum_{i=1}^{M} \frac{|\text{DB}_i|}{\text{hd}_i^2}} \qquad \text{F}_{ecn} = \frac{2*\text{Precision}_{ecn}*\text{Recall}_{ecn}}{\text{Precision}_{ecn}+\text{Recall}_{ecn}}$$

where GB_i is the ground truth bounding box number i and hg_i is its height, while DB_i is the detected bounding box number i and hd_i is its height. N is the number of ground truth bounding boxes and M is the number of detected bounding boxes and GDI, DGI are the corresponding intersections:

$$\text{GDI}_i = \text{GB}_i \cap \left(\bigcup_{i=1}^{M} \text{DB}_i \right) \qquad \text{DGI}_i = \text{DB}_i \cap \left(\bigcup_{i=1}^{N} \text{GB}_i \right)$$

6.1 Evaluation based exclusively on character pixels

The proposed algorithm [48] (Fig.7) generates two binarized images for the ground truth and the resulted bounding boxes respectively. For each case, the algorithm takes as input the image and a set of

bounding boxes. The pixels contained in each box are binarized using Otsu thresholding [36] and then conditionally inverted producing black pixels for text and white pixels for the background. All pixels outside boxes are set to white. Otsu's method proved to be a very good solution for this kind of text images although the choice of the binarization method does not affect considerably the result of the proposed evaluation because of the skeletonizing that follows.
Then, the two binarized images have to be compared in order to compute the recall and precision rates. However, a straight comparison between the two images would produce evaluation rates strongly depended to the used binarization method. To overcome this problem and focus only on the evaluation of text detection we use the skeleton of each image computed by an iterative skeletonization method presented in [49]. Specifically, the recall and precision rates are defined by the equations:

$$\text{Recall} = \frac{|\text{GT\_skel} \cap \text{RS\_bin}|}{|\text{GT\_skel}|} \qquad \text{Precision} = \frac{|\text{RS\_skel} \cap \text{GT\_bin}|}{|\text{RS\_skel}|}$$

where GT_bin is the binarized ground truth image and GT_skel is its skeleton while RS_bin is the binarized image of the detection result and RS_skel is its skeleton. The operator |…| denotes the number of text (black) pixels and ∩ is the intersection of the text (black) pixels.

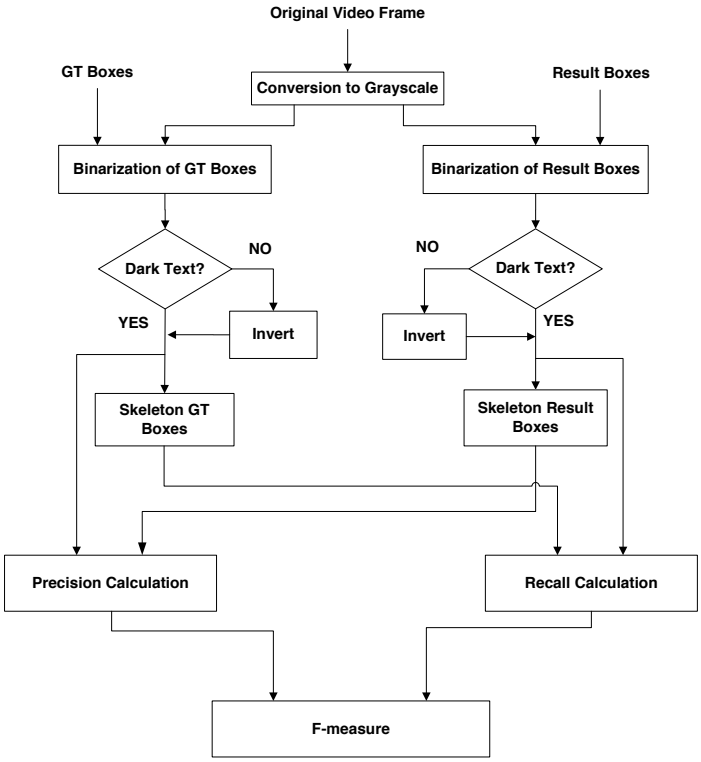


Figure 7. Flowchart of the proposed algorithm

7 Experimental results

For the experiments regarding the text detection methods we created two datasets. The first one consists of 214 frames from athletic videos while the second contains 172 video frames from news broadcasts. Table 1 shows comparative results for the two proposed text detection methods [45, 47] against three state-of-the-art algorithms. The used evaluation protocol is the one presented in section 6.1. Table 2 presents the average processing times for the same methods. Table 3 present the results of the proposed method [47] for the dataset of the competition Robust Reading of ICDAR 2003 and 2005 compared to the other methods of the competition. These experiments combined with the ones of table 1, proved the capability of the proposed method to deal with both artificial and scene text in camera-based images or video frames. Table 4 provides a feature set comparison in terms of classification accuracy combined with two classifiers, SVM and Random Forest. According to these results the best performance is achieved by reduced eLBP feature set combined with an SVM classifier as proposed in [45]. However, the difference in terms of performance between SVM and RF best results is negligible compared to the corresponding computational cost as it is shown in table 5. This exceptional efficiency

of Random Forest allowed us to create a pure machine learning system [47] which outperformed any other method in literature.

Table 1. Comparative results for artificial text detection algorithms

%	Method	$Recall_{ecn}$	$Precision_{ecn}$	F_{ecn}
Dataset 4	[20]	66.9	66.7	66.8
	[13]	65.4	75.6	70.1
	[23]	81.1	70.5	75.4
	Proposed [45]	83.9	79	81.4
	Proposed [47]	84.1	79.8	81.9
Dataset 5	[20]	63.3	69.2	66.1
	[13]	68.2	71.1	69.6
	[23]	80.6	71.5	75.8
	Proposed [45]	82.7	83.5	83
	Proposed [47]	83.4	84.7	84

Table 2. Average processing time per frame for *Dataset 4* and *Dataset 5*

Method	Average Processing time per frame (secs)
[20]	8
[13]	3.35
[23]	1.5
Proposed [45]	2
Proposed [47]	1.6

Table 3. Comparative results for ICDAR2003 dataset and evaluation protocol

	Method	*p*	*r*	*f*	T(sec)
	Proposed [47]	0.82	0.61	0.70	4.2
	Ji [19]	0.59	0.79	0.68	-
	Epshtein [8]	0.73	0.60	0.66	0.94
Robust Reading Competition 2005	**Hinnerk Becker**	0.62	0.67	0.62	14.4
	Alex Chen	0.60	0.60	0.58	0.35
	Qiang Zhu	0.33	0.40	0.33	1.6
	Jisoo Kim	0.22	0.28	0.22	2.2
	Nobuo Ezaki	0.18	0.36	0.22	2.8
Robust Reading Competition 2003	**Ashida**	0.55	0.46	0.50	8.7
	HWDavid	0.44	0.46	0.45	0.3
	Wolf	0.30	0.44	0.35	17
	Todoran	0.19	0.18	0.18	0.3

Table 4. Classification results of *Dataset 1* for different features and classifiers

	Features	Feature dimension	$Recall_{text}$	$Precision_{text}$	F_{text}
RF	MACeLBP	2048	96.6	97.4	**97**
	Reduced eLBP	256	93.5	94.6	**94.1**
	DCT	287	94.2	91.1	**92.7**
	Haar	279	91.4	90	**90.1**
	Gradient	288	90.5	88.2	**89.3**
SVM	Reduced eLBP	256	98	98	**98**
	MACeLBP	2048	96	98.2	**97.1**
	DCT	287	94.2	96.2	**95.2**
	Haar	279	93.1	95.7	**94.4**
	Gradient	288	80.1	92.3	**86.3**

Table 5. Comparing SVM and RF in terms of prediction speed.

(Predictions/sec)	SVM	RF
MACeLBP	40	~150.000
Reduced eLBP	160	~500.000
DCT	180	~500.000
Haar	385	~500.000
Gradient	200	~500.000

8 Conclusion

In this thesis, novel methodologies for the extraction of textual information from images and videos have been proposed. Three different algorithms for text detection were developed covering all cases of text in all kinds of media. The first is a very efficient edge-based heuristic algorithm for the detection of artificial text in videos. The second one is a two-stage, coarse to fine algorithm which uses the previous approach as a first step while also integrates a second machine learning stage that refines the result. The success of this stage is based on a very powerful feature set, in terms of discrimination which is produced by a new operator called edge Local Binary Pattern (eLBP). The eLBP feature set describes the distribution of local edge patterns which actually distinguishes text from the background. The third proposed text detection method uses directly machine learning to locate text. The feature set is based on Multilevel Adaptive Color eLBP (MACeLBP) which constitutes an evolved eLBP operator that includes additionally color edge information and operates edge multilevel thresholding in a locally adaptive way. The use of the very efficient Random Forest classifier gave us the capability to scan with a sliding window the whole image in several multiresolution levels and develop a completely parameter-free machine learning system which is able to detect both artificial and scene text in camera-based images or video frames. Two text detection evaluation methods were also proposed. Both of them are based on intuitively correct criteria and try to measure objectively the percentage of the retrieved textual information. The first evaluation protocol computes recall and precision rates based on the estimated character number of each text bounding box. The second algorithm passes the evaluation process to the next stage of text binarization, computing the retrieval rates based exclusively on character pixels. In that way we overcome all the problems caused by the ambiguously defined text bounding boxes. In this framework we developed a text pixel segmentation system which binarizes the image and then detects and inverts cases of inverse text with a very high accuracy. Finally, a very efficient method was developed for the temporal detection of text which actually conduces to a better spatial detection while concurrently enhances the quality of the text image.

References

1. Lienhart R and Effelsberg W (2000) Automatic Text Segmentation and Text Recognition for Video Indexing. ACM/Springer Multimedia Systems, Vol. 8. pp.69-81
2. Sobottka K , Bunke H, Kronenberg H (1999) Identification of Text on Colored Book and Journal Covers. International Conference on Document Analysis and Recognition, pp. 57–63.
3. Wang K, Kangas J.A (2003) Character Location in Scene Images from Digital Camera. Pattern Recognition, Volume 36, Number 10, pp. 2287-2299(13)
4. Sato T, Kanade T, Hughes E, and Smith M (1998) Video OCR for Digital News Archives, IEEE Workshop on Content-Based Access of Image and Video Databases , pp. 52 – 60
5. Kim W, Kim C (2009) A New Approach for Overlay Text Detection and Extraction from Complex Video Scene. IEEE Transactions on Image Processing, vol.18, no.2, pp.401-411
6. Chen X, Yang J, Zhang J, Waibel A (2004) Automatic Detection and Recognition of Signs from Natural Scenes, IEEE Transactions on Image Processing, Vol. 13, No. 1. pp. 87-99.
7. Epshtein B, Ofek E, Wexler Y (2010) Detecting Text in Natural Scenes with Stroke Width Transform, IEEE Conference on Computer Vision and Pattern Recognition, San Francisco.
8. Zhong Y, Zhang H and Jain A.K (2000) Automatic Caption Localization in Compressed Video. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(4): pp.385–392.
9. Crandall D, Antani S, Kasturi R (2003) Extraction of Special Effects Caption Text Events from Digital Video. International Journal on Document Analysis and Recognition (5), No. 2-3, pp. 138-157
10. Lim Y.K, Choi S.H, and Lee S.W (2000) Text Extraction in MPEG Compressed Video for Content-based Indexing. International Conference on Pattern Recognition, pp. 409-412.
11. Gargi U, Crandall D.J, Antani S, Gandhi T, Keener R, Kasturi R (1999) A System for Automatic Text Detection in Video. International Conference on Document Analysis and Recognition, pp. 29-32
12. Goto H (2008) Redefining the DCT-based feature for scene text detection: Analysis and comparison of spatial frequency-based features. International Journal on Document Analysis and Recognition (11), No. 1, October 2008, pp. 1-8.
13. Chen D, Odobez J-M and Thiran J-P (2004) A Localization/Verification Scheme for Finding Text in Images and Videos Based on Contrast Independent Features and Machine Learning Methods. Image Communication, vol. 19(3), pp. 205-217
14. Ye Q, Huang Q, Gao W, Zhao D (2005) Fast and Robust Text Detection in Images and Video Frames. Image Vision Computing, 23(6): pp.565-576.
15. Jung C, Liu Q, Kim J (2009) A Stroke Filter and its Application to Text Localization. Pattern Recognition Letters, 30(2): pp. 114-122.
16. Ye Q, Jiao J, Huang J, Yu H (2007) Text detection and restoration in natural scene images, Journal of Visual Communication and Image Representation 18(6), pp. 504-513.
17. Ji R, Xu P, Yao H, Zhang Z, Sun X, Liu T (2008) Directional correlation analysis of local Haar binary pattern for text detection. IEEE International Conference on Multimedia & Expo, pp.885-888.
18. A. Ekin. Information Based Overlaid Text Detection by Classifier Fusion. IEEE International Conference on Acoustics, Speech and Signal Processing, (2006), pp. II-753-756
19. Jung K (2001) Neural Network-based Text Location in Color Images. Pattern Recognition Letters , 22(14): pp. 1503–1515.
20. Kim K.I, Jung K, Park S.H and Kim H.J (2001) Support Vector Machine-based Text Detection in Digital Video. Pattern Recognition, 34(2): pp. 527–529.
21. Wolf C and Jolion J-M (2004) Model Based Text Detection in Images and Videos: a Learning Approach. Technical Report LIRIS-RR-2004-13 Laboratoire d'Informatique en Images et Systemes d'Information, INSA de Lyon, France.
22. Lienhart R and Wernicke A (2002) Localizing and Segmenting Text in Images and Videos. IEEE Trans. on Circuits and Systems for Video Technology, 12(4): pp.256–268.
23. Li H, Doermann D and Kia O (2000) Automatic Text Detection and Tracking in Digital Video, IEEE Transactions on Image Processing. Vol. 9, No. 1, pp. 147-156.
24. Rainer Lienhart and Frank Stuber, (1995) "Automatic text recognition in digital videos", Technical Report / Department for Mathematics and Computer Science, University of Mannheim ; TR-1995-036.
25. S. Antani, U. Gargi, D. Crandall, T. Gandhi, and R. Kasturi, "Extraction of Text in Video", Technical Report of Department of Computer Science and Engineering, Penn. State University, CSE-99-016, August 30, 1999.
26. Y. Nakajima, A. Yoneyama, H. Yanagihara, and M. Sugano, "Moving Object Detection from MPEG Coded Data", Proc. of SPIE, 1998, Vol. 3309, pp.988-996.
27. M. Pilu, On Using Raw MPEG Motion Vectors to Determine Global Camera Motion, Proc. of SPIE, 1998, Vol. 3309, pp. 448-459.
28. H. Li and D. Doermann. Text Enhancement In Digital Video Using Multiple Frame Integration. Proceedings of ACM Multimedia 99 , pages 19-22.
29. C.M. Lee and A. Kankanhalli. Automatic extraction of characters in complex scene images. International Journal of Pattern Recognition and Artificial Intelligence, 9(1):67-82, 1995.
30. F. LeBourgeois. Robust Multifont OCR System from Gray Level Images. In Proceedings of the 4th International Conference on Document Analysis and Recognition, pages 1-5, 1997.

31. S. Antani, D. Crandall, and R. Kasturi. Robust Extraction of Text in Video. In Proceedings of the International Conference on Pattern Recognition, volume 1, pages 831-834, 2000.
32. M. Kamel and A. Zhao. Extraction of Binary Character/Graphics Images from Grayscale Document Images. Computer Vision, Graphics, and Image Processing, 55(3):203-217, May 1993.
33. V. Wu, R. Manmatha, E.M. Riseman. Textfinder: An Automatic System to Detect and Recognize Text in Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, Issue 11, pp. 1224-1229, Nov. 1999.
34. D. Chen, K. Shearer, and H. Bourlard. Text enhancement with asymmetric filter for video OCR. In Proceedings of the 11th International Conference on Image Analysis and Processing, pages 192-197, 2001.
35. D. Chen, J.M. Odobez, and H. Bourlard. Text segmentation and recognition in complex background based on markov random field. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, volume 4, pages 227-230, 2002.
36. Otsu N (1979) A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics Vol. 9, No. 1, pp. 62-66.
37. Jun Ohya, Akio Shio, and Shigeru Akamatsu. Recognizing Characters in Scene Images. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 16, No. 2, Febr. 1994.
38. W. Niblack. An Introduction to Digital Image Processing, pages 115-116. Englewood Cliffs, N.J.: Prentice Hall, 1986.
39. J. Sauvola, T. Seppanen, S. Haapakoski, and M. Pietikainen. Adaptive Document Binarization. In International Conference on Document Analysis and Recognition, volume 1, pages 147-152, 1997.
40 V. Manohar, P. Soundararajan, M. Boonstra, H. Raju, D. Goldgof, R. Kasturi, J. Garofolo. Performance Evaluation of Text Detection and Tracking in Video. International Workshop on Document Analysis Systems, pp. 576-587
41 R. Kasturi, D. Goldgof , P. Soundararajan, V. Manohar , J. Garofolo , R. Bowers , M. Boonstra , V. Korzhova , J. Zhang. Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol, IEEE Transactions on Pattern Analysis and Machine Intelligence (2008), 31 (2) pp. 319-336.
42 C. Wolf, J. Jolion. Object Count/area Graphs for the Evaluation of Object Detection and Segmentation Algorithms. International Journal on Document Analysis and Recognition (2006), 8(4) pp. 280-296.
43 M. Anthimopoulos, B. Gatos and I. Pratikakis, "Multiresolution Text Detection in Video Frames", 2nd International Conference on Computer Vision Theory and Applications (VISAPP 2007), pp. 161-166, Barcelona, Spain, March 2007
44 Canny J., 1986. A computational approach to edge detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8, 679-698.
45 M. Anthimopoulos, B. Gatos, I. Pratikakis, «A two-stage scheme for text detection in video images», Image and Vision Computing, Vol. 28 , Issue 9, pp. 1413-1426, 2010.
46 M. Anthimopoulos, B. Gatos, and I. Pratikakis, "A Hybrid System for Text Detection in Video Frames", 8th International Workshop on Document Analysis Systems (DAS'08), pp. 286-292, Nara, Japan, September 2008.
47 M. Anthimopoulos, B. Gatos, I. Pratikakis, “Detection of Artificial and Scene Text in Images and Video Frames”, Pattern Analysis and Applications. Accepted for publication.
48 M.Anthimopoulos, N.Vlissidis, B.Gatos, “A Pixel-Based Evaluation Method for Text Detection in Color Images” The 20th International Conference on Pattern Recognition
49. H. J. Lee, B. Chen, "Recognition of Handwritten Chinese Characters via Short Line Segments", Pattern Recognition (1992), 25 (5), pp. 543-552.

Contribution to the modeling of coupled resonator optical waveguides in photonic crystals

Nikolaos Avaritsiotis*

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications

nickava@di.uoa.gr

Abstract. In this dissertation, coupled resonator optical waveguides are analyzed and a new analytical model to study their spectral properties is developed. Coupled mode theory and plane wave expansion are employed in order to study the spectral properties of CROWs and a new semi-analytical model is derived. This semi-analytical model is used for the derivation of an analytical solution for the device's resonant frequencies. The model not only provides a useful tool in the design of CROW based devices but also presents a useful physical insight for the device under investigation. The semi-analytical model is also used in several types of coupled cavity devices, as is a SCISSOR, and is compared with other methods, such as FDTD, in order to test the validity and accuracy of the results. All approximations and assumptions that lead to a simplification of the model are being discussed. Furthermore the relation between fabrication imperfections and the performance of photonic crystal CROWs is examined based on our semi-analytical model. A statistical study of such imperfections is enabled owing to the calculation of the coupling coefficients derivatives. The model is used to study the spectral influence of imperfections of different strengths and types on various coupled cavity devices.

Keywords: Photonic crystal waveguide, CROW, SCISSOR, Coupled mode theory, FDTD, frequency response, geometric perturbations.

1 Dissertation Summary

In this dissertation the technology of photonic crystals is chosen as the technology platform to study optical integrated nanophotonic devices. Photonic crystals are periodic structures created by materials with different dielectric constants that are periodically placed in space. Photonic crystals are classified in 1D, 2D and 3D according to their dielectric constant periodicity in space. One of the most important features of photonic crystals is the presence of photonic band gaps in their dispersion diagrams. More specifically a photonic band gap is a range of frequencies for which light cannot propagate inside the structure. Therefore many useful photonic crystal

*Dissertation Advisor: Thomas Sphicopoulos, Professor.

devices can be constructed with photonic band gaps, preventing light from propagating in certain directions with specified frequencies [1].

Coupled Resonator Optical Waveguide (CROW) devices in photonic crystals may find important applications in future integrated nanophotonic circuits. These devices represent a new kind of waveguides not depending on the principles of total internal reflection nor on Bragg reflectors to guide light [2]. Waveguiding is performed by the coupling of neighboring resonators appropriately placed in order to ensure loose coupling [2]. CROWs can be manufactured by several kinds of resonators such as Fabry-Perot resonators, photonic crystal cavities, micro rings and micro disks and may function with different ways of coupling between these resonators giving rise to several designs and applications. In this dissertation the photonic crystal cavities were chosen to construct CROWs and Side Coupled Integrated Spaced Sequence of Optical Resonators (SCISSORs) [3].

Several methods are going to be developed in order to study these structures. The role of these methods is very important so as to understand the way in which light propagates through these structures and in addition to theoretically analyze and design them. Arithmetic methods that directly solve Maxwell's equations in the time domain (Finite Difference Time Domain, FDTD) [4] or in frequency domain (Finite Difference Frequency Domain, FDFD) [5] can overview the evolution of electromagnetic fields and may be successfully applied to devices like CROWs and SCISSORs. Unfortunately they present several drawbacks such as need for enormous computational power and time for the simulation of complicated structures. The Mode Matching method [6] can analyze the properties of such structures having as only prerequisite the analysis of the structure in supercells. Another similar method that is going to be fully analyzed in this dissertation is the Couple Mode Theory (CMT) [7].

Firstly a closed form formula for the calculation of the transfer function of a Photonic Crystal (PC) CROW coupled to an input and an output PC waveguide will be derived [8]. Coupled Mode Theory is initially used for the derivation of a semi analytical transfer function model [11]. This semi-analytical model is compared to the results of an in-house FDTD tool (also used in [13]) and good agreement is obtained. Using this semi-analytical model and taking into account only adjacent cavity and/or waveguide coupling, a simpler analytical model is obtained for the first time, providing a closed form formula for the transfer function of the device regardless of the number of cavities. This analytical model may be used to quickly estimate the transfer function of the device once the coupling coefficients are estimated. Using the analytical model, the resonant frequencies of the device were also obtained. This model can provide a useful tool in the design of CROW-based filters and other similar devices [8].

As previously stated CROWs and SCISSORs are well suited for coupling of mode analysis, which usually requires much less computational resources compared Finite Difference Time Domain (FDTD) schemes. Coupled mode models [10] also provide a useful physical insight in the device operation. Therefore in this dissertation is derived a general coupled mode theoretic model for the treatment of coupled cavity devices incorporating various phenomena such as dispersion, frequency variation of the coupling coefficients, non-adjacent cavity coupling and waveguide mode self coupling [14]. The model is validated comparing its results against the FDTD method [13] and the strength of the underlying assumptions is highlighted. Overall it was

shown that the CMT model can provide an adequate device description offering a tangible manner of calculating the transfer function and a useful physical insight.

Finally the relation between fabrication imperfections and the performance of a photonic crystal Coupled Resonator Optical Waveguide (CROW) is studied [12]. A semi analytical model is presented, which calculates the perturbation of the coupling coefficients through their derivatives with respect to the geometric characteristics of the rods of the photonic crystal lattice. To account for random perturbations in finite devices, Finite Difference (FD) methods require very small grid size in order to capture small geometric perturbations. To obtain reliable statistical results, many perturbed devices must be calculated rendering such simulations intractable. This alternative approach, based on a previously developed coupled mode model [8], is applied to the calculation of the derivatives of the coupling coefficients with respect to the rod radii and positions. Once these derivatives are calculated, the transfer functions of a large number of devices with randomly perturbed geometric characteristics using Taylor's expansion, are estimated and the results are analyzed discussed.

2 Results and Discussion

2.1 Electromagnetic field equations

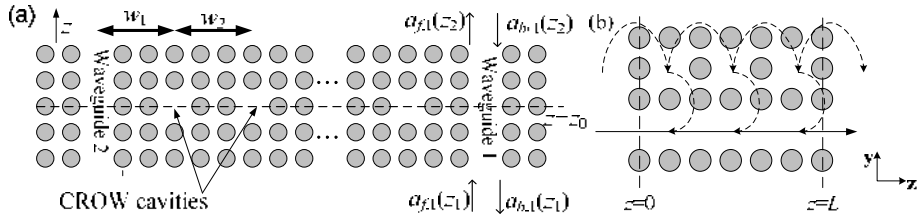


Fig. 1 (a) Coupled Resonator Optical Waveguides (CROW) coupled to two photonic crystal waveguides, and (b) a Side coupled integrated sequence of three resonators (SCISSOR).

The basic equations for the electromagnetic field inside the photonic crystal CROW will be given in order to derive a transfer function model for the structure of Fig. 1a based on CMT. Assuming weak coupling, the electromagnetic field (**E**,**H**) can be accurately approximated as a linear superposition of the isolated modes of the waveguides and cavities, i.e.

$$\mathbf{E} = \sum_n a_n \mathbf{E}_n + \sum_l a_{fl} \mathbf{E}_{fl} + \sum_l a_{bl} \mathbf{E}_{bl} \tag{1}$$

$$\mathbf{H} = \sum_n a_n \mathbf{H}_n + \sum_l a_{fl} \mathbf{H}_{fl} + \sum_l a_{bl} \mathbf{H}_{bl} \tag{2}$$

where $\mathbf{E}_n$, $\mathbf{H}_n$ are the electric and magnetic modal fields of the n^{th} isolated cavity modes ($1 \le n \le N$), ($\mathbf{E}_{fl}$, $\mathbf{H}_{fl}$) and ($\mathbf{E}_{bl}$, $\mathbf{H}_{bl}$) are the forward and backward propagating modes of the l^{th} waveguide (l=1 or 2). Due to the symmetry of the device in Fig. 1a,

the role of the input and output waveguide may be interchanged. The waveguide l=1 will be considered as the input waveguide. In addition a_n, a_{fl} and a_{bl} denote the excitation coefficients of the n^{th} cavity mode and the forward and backward isolated propagating mode of the l^{th} waveguide respectively. Assuming that z is the propagation direction of the waveguide modes, a_{fl} and a_{bl} are generally considered z-dependent [15], while the cavity mode amplitudes a_n are assumed not to depend on z [16]. The Bloch's theorem will be used in order to express the waveguide mode fields as [17]. The propagation constant β_{ml} will be considered positive for the forward (m=f) and negative for the backward (m=b) propagating mode and the Bloch functions e_{ml} and h_{ml} are periodic vector functions with the same periodicity a as the input/output waveguides along the z direction. Since the waveguides are considered the same, at a given frequency ω, the propagation constants will be β_{f1}=β_{f2}=-β_{b1}=-β_{b2}=β. The waveguide modes are normalized and the forward and backward propagating modes obey the orthogonality relations as in [15].

In the same way the cavity modes obey a similar set of equations,

$$\nabla \times \mathbf{E}_n = j\omega_o \mu \mathbf{H}_n \tag{3}$$

$$\nabla \times \mathbf{H}_n = -j\omega_o \varepsilon_{cn} \mathbf{E}_n \tag{4}$$

where $\varepsilon_{cn}(\mathbf{r})$ is the isolated dielectric constant distribution of n^{th} cavity alone and ω_0 is the isolated mode resonant frequency. Assuming a lossless structure, one may chose the electric fields for the cavity modes to be purely real $\mathbf{E}_n^*$=$\mathbf{E}_n$ resulting in purely imaginary magnetic field, $\mathbf{H}_n^*$=-$\mathbf{H}_n$ as discussed in [18].

2.2 Coupled Mode Equations

In order to derive the coupled mode equations for the waveguide and cavity modes of the structure, the Lorentz's reciprocity theorem [18] will be used. This theorem relates two electromagnetic fields ($\mathbf{E}_a$,$\mathbf{H}_a$) and ($\mathbf{E}_b$,$\mathbf{H}_b$) obeying Maxwell's equations in media with dielectric constant distributions ε_a and ε_b respectively. Using the corresponding reciprocity equations presented in [18], the following coupled system of equations is derived:

$$\frac{d\mathbf{a}}{dz} = \mathbf{W}\mathbf{a} + \mathbf{C}\mathbf{b} \tag{5}$$

where the vectors $\mathbf{a}$=(a_1,a_2) and $\mathbf{b}$=(a_{c1},a_{c2},…,a_{cN}) contain the amplitudes of the waveguide and cavity mode respectively, while the matrices $\mathbf{W}$=[w_{pq}] and $\mathbf{C}_{pq}$=[c_{pq}] are determined by the coupling coefficients of the modes,

$$w_{pq} = j\omega\delta_p \int_S dS\left(\varepsilon - \varepsilon_w\right)\mathbf{E}_p \cdot \mathbf{E}_q^* \tag{6}$$

$$c_{pq} = j\delta_p \left\{ \mu\left(\omega - \omega_0\right)\int_S dS\mathbf{H}_{cq} \cdot \mathbf{H}_p^* + \int_S dS\left(\varepsilon\omega - \varepsilon_{cp}\omega_0\right)\mathbf{E}_{cq} \cdot \mathbf{E}_p^* \right\} \tag{7}$$

where δ_1=1 and δ_2=-1. The coefficients w_{pq} correspond to the coupling of the waveguide modes inside the cavities, while c_{pq} are determined by the coupling of the waveguide modes with the cavity modes. As previously stated we apply Bloch's theorem [1] to express the forward waveguide mode as $\mathbf{E}_1=\mathbf{e}_1\exp(j\beta z)$, where β is the propagation constant of the mode and $\mathbf{e}_1$ is the Bloch function which is periodic along z. In order to obtain the cavity coupled mode equations, the reciprocity equations can be applied in this case considering the vector functions $\mathbf{F}_n=\mathbf{E}\times\mathbf{H}_{cn}^{*}+\mathbf{E}_{cn}^{*}\times\mathbf{H}$, where $1\le n\le N$.

The following equations are then derived:

$$\int_0^L dz\mathbf{Sa}+\mathbf{Kb}=0 \tag{8}$$

In (8) the integration is performed along the propagation direction from the input (z=0) to the output (z=L) of the device. The matrices $\mathbf{S}=[s_{pq}]$ and $\mathbf{K}=[\kappa_{pq}]$ are defined by:

$$s_{pq}=j\mu(\omega-\omega_0)\int_S dS\mathbf{H}_{cp}^{*}\cdot\mathbf{H}_q+j\int_S dS\left(\varepsilon\omega-\varepsilon_{cp}\omega_0\right)\mathbf{E}_{cp}^{*}\cdot\mathbf{E}_q \tag{9}$$

$$\kappa_{pq}=j\mu(\omega-\omega_0)\int_V dV\mathbf{H}_{cq}\mathbf{H}_{cp}^{*}+j\int_V dV\left(\varepsilon\omega-\varepsilon_{cq}\omega_0\right)\mathbf{E}_{cq}\mathbf{E}_{cp}^{*} \tag{10}$$

where elements κ_{pq} are the coupling coefficients between the cavity modes and s_{pq} much like c_{pq} are determined by the coupling of the waveguide modes with the cavity modes.

2.3 Estimation of the transfer function

The previous equations provide a framework for the estimation of the transfer function of the structure. The power transfer function is defined as $T(\omega)=|a_1(L)/a_1(0)|^2$ and is determined by the ratio of the amplitudes of the forward propagating mode at the device output and input. In [8] we have shown how under certain simplifying assumptions (e.g. assuming $\mathbf{W}\cong 0$) the transfer function of the CROW can be obtained. However in this case a more generalized transfer function derivation will be shown for the case of a SCISSOR (Fig. 1b). Because of the term corresponding to **W**, (5) is not directly amenable to integration as in [8] and for this reason we consider the 2×2 matrix **U**, obeying the differential equation:

$$\frac{\partial\mathbf{U}}{\partial z}=\mathbf{WU} \tag{11}$$

Given its value **U**(0) at z=0, **U** can be calculated numerically by approximating the derivative in (11) with a finite difference. If **W** is small enough, then it can be easily shown that **U** is approximated by:

$$\mathbf{U}(z) \cong \mathbf{I} + \int_0^z dz' \mathbf{W}(z') \tag{12}$$

where it is assumed that $\mathbf{U}(0)=\mathbf{I}$. We substitute $\mathbf{a}=\mathbf{U}\mathbf{c}$ in (5) in which case we obtain:

$$\frac{d\mathbf{c}}{dz} = \mathbf{U}^{-1}\mathbf{C}\mathbf{b} \tag{13}$$

The above equation can now be readily integrated with respect to z, in order to obtain:

$$\mathbf{a}(L) = \mathbf{U}^{-1}(L)\left\{\mathbf{a}(0) + \left(\int_0^L dz \mathbf{U}^{-1}\mathbf{C}\right)\mathbf{b}\right\} \tag{14}$$

where we used the fact that $\mathbf{c}=\mathbf{U}^{-1}\mathbf{a}$. The vector $\mathbf{b}$ can be estimated by (8), if one performs integration by parts. We assume a matrix $\mathbf{\Lambda}$ such that:

$$\mathbf{\Lambda}(z) = \int_0^z dz' \mathbf{S}(z')\mathbf{U}(z') + \mathbf{P} \tag{15}$$

where $\mathbf{P}$ is a constant matrix. Taking into account that $\mathbf{S}\mathbf{a}=(\partial\mathbf{\Lambda}/\partial z)\mathbf{c}$, we can write:

$$\int_0^L dz \mathbf{S}\mathbf{a} = \mathbf{\Lambda}(L)\mathbf{c}(L) - \mathbf{\Lambda}(0)\mathbf{c}(0) - \int_0^L dz \mathbf{\Lambda}\frac{d\mathbf{c}}{dz} \tag{16}$$

and using (13) and (8) we obtain:

$$\mathbf{\Lambda}(L)\mathbf{U}^{-1}(L)\mathbf{a}(L) - \mathbf{\Lambda}(0)\mathbf{U}^{-1}(0)\mathbf{a}(0) - \mathbf{G}\mathbf{b} = 0 \tag{17}$$

where matrix $\mathbf{G}$ is determined by:

$$\mathbf{G} = -\mathbf{K} + \int_0^L dz \mathbf{\Lambda}\mathbf{U}^{-1}\mathbf{C} \tag{18}$$

We note that in (17), the amplitudes of the cavity modes contained in $\mathbf{b}$ are expressed in terms of the input and output waveguide mode amplitudes $\mathbf{a}(L)$ and $\mathbf{a}(0)$. The amplitudes $a_1(0)$ and $a_2(L)$ of the forward and backward mode at $z=0$ and $z=L$ are determined by the incident wave conditions. Typically when calculating the transfer function, we assume that $a_2(L)=0$, i.e. that there is no reflected wave at the device output. In any case, we can choose the elements of $\mathbf{P}$, so that only the incident amplitudes of the forward and backward modes $a_1(0)$ and $a_2(L)$ respectively appear in (17). To do this we require the first column of $\mathbf{\Lambda}(L)\mathbf{U}^{-1}(L)$ to be zero and that the second column of $\mathbf{\Lambda}(0)\mathbf{U}^{-1}(0)$ be also zero. Using some straightforward mathematic manipulations, we find that the elements of P_{pq} of $\mathbf{P}$ must be given by:

$$P_{p2} = 0 \tag{19}$$

$$P_{p1} = \int_0^L dz \left(\frac{u_{21}(L)}{u_{22}(L)} M_{p2}(L) - M_{p1}(z) \right) \tag{20}$$

where the $\mathbf{M}=[M_{pq}]$ is the matrix $\mathbf{M}=\mathbf{SU}$ and u_{pq} are the elements of $\mathbf{U}$ which can be estimated numerically using (11) or (12). Given M_{pq}, we can use (19)-(20) to determine the elements of $\mathbf{P}$. Assuming that $a_2(L)=0$, then taking into account that $\mathbf{E}(0)=\mathbf{I}$ and that $\mathbf{\Lambda}(0)=\mathbf{P}$, we obtain from (17):

$$\mathbf{b} = a_1(0)\mathbf{G}^{-1}\mathbf{P}^{(1)} \tag{21}$$

where $\mathbf{P}^{(q)}$ denotes the q^{th} column of $\mathbf{P}$. Equation (21) expresses the cavity mode amplitudes $\mathbf{b}=(a_{c1},a_{c2},\ldots,a_{cN})$ in terms of the forward waveguide mode amplitude $a_1(0)$ at the device input. Therefore equation (14) can be used to estimate $a_1(L)$ in terms of $a_1(0)$ and $\mathbf{b}$. The transfer function can be calculated taking into account the fact that $T(\omega)=|a_1(L)/a_1(0)|^2$.

To calculate the transfer function of a CROW, like the one depicted in Fig. 1a, using the semi-analytical form and the simplifying assumptions of [8] one first needs to estimate the coupling coefficients of the cavity/cavity and cavity/waveguide systems and hence the isolated modal fields of the waveguides and the cavities. This can be achieved through the Plane Wave Expansion (PWE) method [17]. For the calculation of the modes of the isolated cavity the number of plane waves used were 55 in each direction (resulting in a total number of 3025 plane waves) while for the waveguide modes the number of plane waves used were 15 along the propagation direction and 61 in the transverse direction. Using this method, the isolated cavity mode resonant frequency f_0 was calculated near $af_o/c=0.3869$. Fig. 2 depicts the power transfer function $T=|H|^2$, obtained. The three notches of T are due to the resonances of the three cavity system (Fig. 2). The amplitude of the power transfer function reaches approximately the value 0.25 at the resonant frequencies [19].

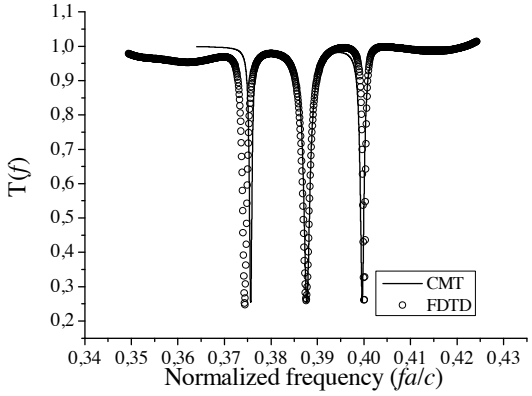


Fig. 2 Transfer functions of a CROW side-coupled to two PC waveguides as obtained using either the FDTD method or the outlined CMT.

In a similar way Fig. 3 presents the results for the generalized CMT model for a SCISSOR (Fig. 1b) along with the FDTD transfer functions for validation purposes.

A single cavity SCISSOR is chosen (depicted in the inset of Fig. 3) device and various grid sizes Δ were assumed for the FDTD scheme. The cavity is spaced one rod away from the waveguide. Fig. 3 illustrates that as grid size reduces the transfer function calculated by the FDTD gradually approaches that of the CMT. For $\Delta=r_a/8$ (r_a is the radius of the rod) one obtains a 0.3% difference between the values of the resonant frequencies predicted by two methods and a 15% difference in the 3 dB bandwidth of the resonance. Smaller grid sizes were not considered because they rendered the FDTD simulations quite time consuming especially in the case of sharper resonances. To estimate the modal fields with the PWE, we used 75x75 plane waves in the case of the cavity mode and 33x75 plane waves in the case of the waveguide mode along the z and y directions respectively. The resonant frequency of the isolated cavity mode was estimated at af_o/c=0.3877. A7x7 and 1x7 supercell was used in the PWE calculation in the case of the cavity and the waveguide fields.

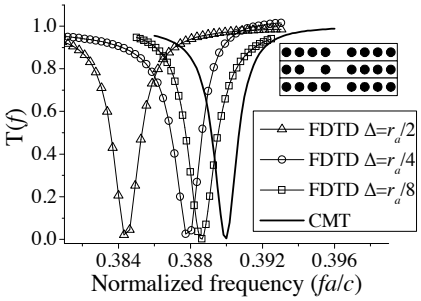


Fig. 3 Transfer function of a single cavity SCISSOR obtained by the CMT and the FDTD scheme. For the latter method, various grid sizes Δ are considered.

2.4 Approximations of the CMT model

The CMT approach usually results in a more tangible estimation of the transfer function of the device. In [8] it was shown how this can be rigorously achieved in the case of a CROW, by ignoring the evanescent waveguide modes in the field expansion, the frequency dependence of the modal fields and secondary coupling effects. In this section some of these restrictions mentioned above will be discussed and the CMT results are going to be compared against the FDTD method.

Frequency dependence of the coupling coefficients.

Inspecting the coupled mode coefficients in (7) and (9)-(10), it is deduced that the coupling coefficients exhibit a frequency dependence since ω appears both in front of the magnetic mode overlap integral and inside the electric field overlap integral. The frequency dependence of the coupling coefficients is also due to the fact that the modal fields are given by $\mathbf{E}_1=\mathbf{e}_1\exp(j\beta z)$ [8] where both the propagation constant β and the Bloch field $\mathbf{e}_1$ are frequency dependent. Our simulation results have shown that taking into account the frequency dependence of the coupling coefficients can have an important bearing on the results.

Coupling Assumptions.

As previously stated, the general coupling of modes analysis accounts for the coupling of the waveguide modes inside the cavities through the matrix **W**. If this secondary coupling is ignored (**W**≅0) the transfer function evaluations become much simpler. It is therefore interesting to investigate the influence of this waveguide mode coupling. The simulation results imply that **W** has a rather minor bearing even in the case where the cavities are placed relatively near the waveguides. Its influence may be greater in the case of structures with weaker mode confinement, however.

In addition it is interesting to consider whether adjacent cavity coupling alone is sufficient to provide an accurate estimate for the transfer function. In this case, the matrix **K** is considered tridiagonal, i.e. $\kappa_{pq}\cong 0$ when $|p-q|>1$. The simulation results indicate that coupling between non-adjacent cavities also results in a frequency detuning which may be important in the case of sharp resonances.

Expansion in terms of the cavity supermodes.

The CMT model presented in 2.2 is based on the expansion of the electromagnetic field in terms of the isolated cavity modes. This is not the only choice however. In [16], we have discussed how an *N*-cavity system can be considered as a single resonator exhibiting *N* modes inside the bandgap, which can be referred as the "supermodes" $\mathbf{E}_{sn}$, $\mathbf{H}_{sn}$ of the cavity system in analogy to the supermodes of a waveguide coupler [20]. These modes obey Maxwell's equations, e.g. $\nabla\times\mathbf{H}_{sn}=j\omega_n\varepsilon_c\mathbf{E}_{sn}$ where ε_c is the dielectric constant of the *N*-cavity system, and ω_n is the resonant frequency of the n^{th} mode. One can apply the reciprocity relations [8] again to obtain coupled mode equations similar to 2.2. The simulation results have shown that the choice of supermodes seems to produce a more accurate description for the broad resonance. However, since it involves the estimation of the cavity modes of a large resonator (coupled cavity system) such estimations may require an excessive number of plane waves.

Evanescent waves.

The simulation results have indicated that in specific cases CMT fails to provide an accurate description of very sharp resonances. This may be due to the assumptions made during its derivation. Probably the most important one is that evanescent waves are neglected in the field expansion of (1)-(2). Although coupled mode theory could in principle be expanded to include evanescent waves [21], this would lead to a cumbersome model, since the number of evanescent waves is infinite. Evanescent modes are included in FDTD, but their influence cannot be easily distinguished. To obtain some measure of the importance of evanescent modes, we resort to the Mode Matching (MM) method also developed in [6] where the PWE method was adapted to estimate evanescent waveguide modes as well.

2.5 Geometric perturbations

To account for random perturbations in finite devices a very small grid size is required in order to capture these small geometric perturbations. To obtain reliable

statistical results, many perturbed devices must be calculated. Therefore an alternative approach is proposed derived from the previously developed coupled mode model based on the calculation of the derivatives of the coupling coefficients with respect to the rod radii and positions [14]. Once these derivatives are calculated, one may estimate the transfer functions of a large number of devices with randomly perturbed geometric characteristics using Taylor's expansion. Consequently the first step is to calculate all the coupling coefficients of the ideal device (i.e. a device free of geometric perturbations) and estimate its transfer function. To incorporate the effect of perturbations, the derivatives of the coupling coefficients need to be calculated as in [14]. Then one may generate random perturbations along the horizontal (Δx_v) and vertical (Δz_v) axis and the rod radius ΔR_v for each rod of each perturbed device. The derivatives in this method need to be calculated only once and can then be used to statistically study the effect of imperfections on a large number (for example 1000) of perturbed devices. The transfer functions for perturbed CROW devices with 10 cavities, for different cavity/waveguide and cavity/cavity spacing are shown in Fig. 4.

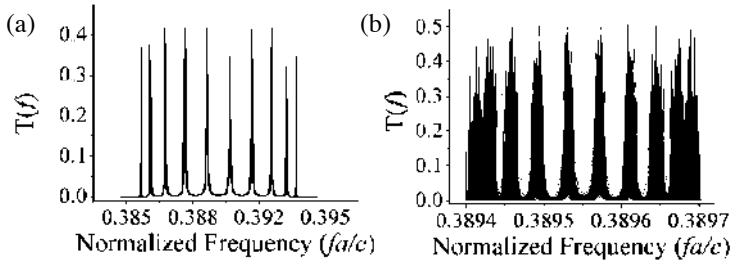


Fig. 4 Transfer functions for perturbed CROW devices with 10 cavities and (a) 2 rod cavity/cavity and waveguide/cavity spacing and (b) three rod spacing. The rod positions and radii are perturbed by 2nm.

Finally simulation results calculated from 200 sample devices with Δ=2nm, depicted in Fig. 5, have shown that the average of standard deviation σ_i=std$\{T_i(f_{ni})\}$, of the resonance centers f_{ni} increases with the size of the device, implying that larger CROWs are much more susceptible to fabrication imperfections.

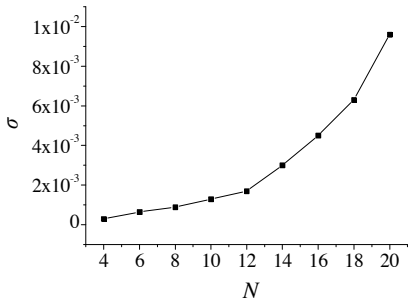


Fig. 5 The average standard deviation σ with respect to the number of CROW cavities for perturbations of Δ=2nm

3 Conclusions

This dissertation presented the method that was developed in order to analyze the spectral characteristics of CROW waveguides with finite cavity numbers in photonic crystals starting from Maxwell's equations. The analysis additionally assumed an input and output waveguide coupled to the CROW structure since its spectral behavior is directly linked to the way light is coupled in and out of the structure. The simplification of the model led to an analytical equation for the calculation of the resonant frequencies of the structure directly by the calculation of its coupling coefficients. The presented method does not require excessive computational time or resources in order to produce accurate results. In addition the results were compared with the FDTD method and good agreement was observed. Furthermore this analytical method provides a better physical insight of the way cavities and waveguides interact inside the CROW and may provide a useful tool in the design of CROW-based filters and other similar devices.

An analysis of the spectral characteristics of SCISSOR devices was also performed using the proposed semi-analytical model. All parameters (i.e. frequency dependence of the coupling coefficients, adjacent cavity/cavity and cavity/waveguide coupling, CMT expansion consideration, evanescent waves) involved in the calculation of the transfer function were discussed and a better physical understanding of the analysis was provided. Many discrepancies of the CMT model for specific SCISSOR structures were therefore explained.

Finally, the influence of fabrication induced disorders in the performance of a photonic crystal CROW was numerically investigated. The semi-analytical model was also used for the calculation of the transfer function of the device in the presence of geometric perturbations of various types. The model was based on the estimation of the derivatives of the coupling coefficients of the CROW. Then, using these derivatives a large number of perturbed devices was simulated with a very small computational overhead. The statistical study of various performance issues such as the amplitude change and frequency shift of the device resonances as well as the resonance 3dB bandwidth were discussed. Using this model one may numerically estimate the relation between the device performance and the quality of the fabrication process.

4 References

1. J.D. Joannopoulos, S.G. Johnson, J.N. Winn, R.D. Meade, Photonic Crystals: Molding the Flow of Light, Princeton University Press, New Jersey, 2008.
2. A. Yariv, Y. Xu, R.K. Lee and A. Scherer, "Coupled-resonator optical waveguide: a proposal and analysis," Optics Letters, Vol. 24, No. 11, pp. 711 713, June 1999.
3. J.E. Heebner, R.W. Boyd, "'Slow' and 'fast' light in resonator-coupled waveguides" J. Modern Opt., Vol. 49, Nov.2002, pp. 2629-2636.
4. A. Tafflove, S. Hagness Computational Electrodynamics: the finite difference time-domain method, Artech House Publishers, 2000.

5. Chin-ping Yu and Hung-chun Chang, "Compact finite-difference frequency-domain method for the analysis of two-dimensional photonic crystals," *Opt. Express* 12, 1397-1408 (2004).
6. A. Theocharidis, T. Kamalakis, and T. Sphicopoulos, "Analysis of photonic crystal waveguide discontinuities using the mode matching method and application to device performance evaluation", JOSA B 24, 1698-1706 (2007).
7. H. A. Haus, Waves and Fields in Optoelectronics, Prentice-Hall 1984.
8. N. Avaritsiotis, T. Kamalakis, T. Sphicopoulos, "Analytical and Numerical Treatment of the Spectral Properties of a Photonic Crystal Coupled Resonator Optical Waveguide" *IEEE J. Lightw. Technol.* To be published.
9. A. Yamilov and M. Bertino, "Disorder-immune coupled resonator optical waveguide" Opt. Lett., Vol. 32, Issue 3, pp. 283-285.
10. S. D. Wu and E. N. Glytsis, "Finite-number-of-periods holographic gratings with finite-width incident beams: analysis using the finite-difference frequency-domain method", *J. Opt. Soc. Am. A*, Vol. 19, No. 10, October 2002, pp. 2018.
11. H. A. Haus and W. P. Huang, "Mode coupling in tampered structures", *IEEE J. Lightw. Technol.*, vol. 7, no. 4, pp. 729-730, Apr. 1989.
12. Avaritsiotis, N.; Kamalakis, T.; Sphicopoulos, T.; , "On the Effectiveness of Coupled Mode Theory in the Analysis of Photonic Crystal Coupled Resonator Devices," *Lightwave Technology, Journal of* , vol.29, no.5, pp.736-743, March1, 2011
13. T. Kamalakis and T. Sphicopoulos, "Numerical study of the implications of size nonuniformities in the performance of photonic crystal couplers using coupled mode theory", *IEEE J. Quantum Electron.*, vol. 41, no. 6, pp. 863-871, June 2005.
14. Avaritsiotis, N.; Kamalakis, T.; Sphicopoulos, T.; , "A Semi-Analytical Model for Numerical Study of a Photonic Crystal Coupled Resonator Optical Waveguide With Disorder," *Lightwave Technology, Journal of* , vol.27, no.14, pp.2892-2899, July15, 2009
15. A. W. Snyder and J. D. Love, *Optical Waveguide Theory*. New York: Chapman and Hall, 1983.
16. T. Kamalakis and T. Sphicopoulos, "Analytical expressions for the resonant frequencies and modal fields of finite coupled cavity chains", *IEEE J. Quantum Electron.*, vol. 41, no. 11, pp. 1419-1425, November 2005.
17. K. Sakoda, *Optical Properties of Photonic Crystals*. Berlin, Germany: Springer-Verlag, 2001.
18. R. E. Collin, *Field Theory of Guided Waves*, ser. Electromagnetic Wave Theory, 2nd ed. Piscataway, NJ: IEEE Press, 1990.
19. C. Manolatou, M.J. Khan, S. Fan, P.R. Villeneuve, H.A. Haus, J.D. Joannopoulos, "Coupling of modes analysis of resonant channel add-drop filters", *IEEE J. Quantum Electron.* Vol. 35, Issue: 9, pp. 1322-1331, Sep 1999.
20. Silberberg, Y., Stegeman, G.I. "Nonlinear coupling of waveguide modes", *Applied Physics Letters*, 50 (13), pp. 801-803 (1987).
21. D. Marcuse, *Theory of Dielectric Optical Waveguides*, Academic Press,2nd edition (1991).

Techno-economic Evaluation of Telecommunications Networks and Services

Konstantina Deligiorgi*

Department of Informatics and Telecommunications, University of Athens
Panepistimiopolis, Ilissia, Athens, Greece, GR-15784
ntina@di.uoa.gr

***Abstract*—**This doctoral thesis presents a methodological framework which integrates, in a uniform way, a number of important parameters that influence the prices of telecommunications products and services. Given the above methodological framework, significant problems of telecommunication market regarding price indexes, are faced. The approaches developed are based on an appropriate mathematical and statistical background and they are applied to specific case studies, providing highly accurate results. The objective of this thesis corresponds to an important part of the techno-economic design aiming to the construction of price indices in the telecommunications market. The design of networks together with the preferences of the customers, are important elements related to the development of the corresponding infrastructure. Due to the rapidly developing technologies and the growing demand for access, design of these networks should provide and support innovative network services and technologies. Determination of a number of parameters, such as the physical characteristics of the telecommunication products and services, the profile of users, the number of users, the socioeconomic factors that influence the telecommunication sector and the shaped market shares due to competition, should be the drivers for the pricing policy of network services and technologies and the development of the infrastructure to support the network operation.

***Index Terms*—**Techno- economic Evaluation, Matched Models, Hedonic Price Indices, Leased Lines, ADSL connections, Physical and Socioeconomic characteristics

I. INTRODUCTION

While price indexing has been a common feature of many products, exactly what constitutes price indexing has been the subject of some disagreement. Price index of products is a research field facing a high level of interest and it is applied for all new and innovative products [1, 2], [4], [5], [6-8]. In case that, false price indices have been constructed, then the consequences for the product marketing will probably be dramatic, since they implicit lack of meeting the customers' preferences and the market's competition leading to oversupply and unneeded investments. The most appropriate case to be considered as an example is the telecommunication sector, since it is almost always connected to significant contemporary investments, regarding new technologies and services and critical business plans, targeting to meet the customers' needs and the market's competition.

As telecommunication services and products are being improved and developed at a fast pace especially during the last ten years, industrial plans rolled out in an attempt to attract and to retain customers. For this purpose, they must precisely know what influence or determines the products' prices, how prices for products that enter the market for the first time or have been modified can be estimated and finally the determination of a price index for these products in a specific period, m, giving the trend of products' prices over time.

* Dissertation Advisor: Thomas Sphicopoulos, Professor

Although literature regarding price indices for established products and services is well developed, new opportunities have emerged due to the nature of high technology products' market. Therefore, further methodological work should be done, by identifying the gaps that have opened up in pricing policy, due to the rapid change of the products' and services' prices and the obvious markets' behavior to fulfill the customers' needs and preferences..

II. THEORETICAL BACKGROUND

One of the main central themes is the mathematical modeling of price indexing, for different types of products and under different assumptions. So far the established research has examined two possible price indexing schemes, under the umbrella of econometric methods. Econometric methods have been used for price index calculation for a long period of time: cars [2], refrigerators [5] and computers [4] are some examples. Furthermore indices about information technology can be found in [1] and [6]. In addition, statisticians use econometric methods in U.S.A but the root of hedonic approach, which is a part of economic research, goes back to [9-12].
There are two types of econometric methods:

- *hedonic methods* and
- *matched model methods*,

each of which have both advantages and disadvantages. One choice is to apply the 'hedonic methods', such as two-period method, single-period method, two-period method with an indicator for new models, or single-regression method. Such indices are commonly used for products, which undergo rapid technological changes.
Hedonic methods refer to regression models in which product prices are related to product characteristics and the observed price of a product (service) is considered to be a function of its characteristics. Generally hedonic methods are based on the idea that a service (product) is a bundle of characteristics and that consumers just buy bundles of product characteristics instead of the product itself. These methods can be used to construct a quality-adjusted price index of a service and researchers described an overview on hedonic price equations [7], [13]. Researcher states that from a large amount of product varieties, consumer chooses without influencing prices [24]. Therefore, consumers maximize utility and producers maximize profits. In hedonic studies it is possible to adjust the price of a service for its quality not quantity. All of them are based on some estimated coefficients that are inflicted on the characteristics of the products in both periods; m and $m+1$. Someone can estimate the coefficients for every year separately or can have observations of two or all years together and estimate a common set of coefficients. The advantage of this method is that calculations are easy and fast. Indeed hedonic methods are very fast to apply but the disadvantage is that index price can change even if no new products are existed, or all prices remain the same.

Another choice is to apply a matched model method such as chained Laspeyres (LCPI), (LPI) or chained Paashe (PCPI), (PPI) or chained Fisher or chained Tornqvist or chained geometric – mean [25]. A classic LPI cannot deal with such complexity due to rapid technological changes or the introduction of new products (services). With LPI an index shows how much the product would cost in period $m+1$ relatively to what it cost in period m. Other price indices function in the same way with slight differences.

The hedonic price indices are commonly used as approximations to the true cost-of-living indices (COLI) which indicate how much money a consumer would need in period $m+1$ relatively to the amount of money he needed in period m so as to keep the same level of utility in period $m+1$ as in period m [3]. The solution is to determine consumer's profile so as to react to a varied and fast-changing supply of products. But how can this profile be determined when everyone has different needs and requirements? No matter what profile is decided, it will be a hypothesis and an

assumption that will respond at a specific model. In addition to the above, someone can see that consumer's desire is not stable and this is not unreasonable because there is a great offer as the 'goods' of technology become more and more attractive. However according to this approach the price index is constructed only using the prices of products, which are available in two adjacent periods.

According to matched model method, Laspeyres in order to create a price index at a time, someone observes the number of units sold in a period m (for example a month) and the average unit price in the period m and m+1. These data are used in the following formula:

$$I_{m+1/m} = \frac{\sum_{i=1}^{n} p_{im+1} q_{im}}{\sum_{i=1}^{n} p_{im} q_{im}} \quad \text{Eq. 1}$$

Price indices are measured, as it is mentioned above, by the matched model method of Laspeyres with chaining average unit prices, which are referred to a previous period, among units sold in the same period.

The term 'hedonic methods' refers to a 'hedonic function' $f(X)$, which is used in economic measurement, where

$$P_i = f(X_i) \quad \text{Eq. 2}$$

where P_i is the price of a variety (or a model) i of a product and X_i is a vector of characteristics associated with the variety. The hedonic function is then used, for different characteristics among varieties of the product, in calculating the price index.

As soon as it is determined which characteristics have to be considered, then the equations (13) and (14) are estimated for N telecommunication products in period m and m+1:

$$\ln(p_{im}) = b_0 + b_1 \cdot X_{1i} + b_2 \cdot X_{2i} + u_{im},\ i=1,\ldots,N \quad \text{Eq. 3}$$
$$\ln(p_{im+1}) = b_0 + b_1 \cdot X_{1i} + b_2 \cdot X_{2i} + b_3 + u_{im+1} \quad \text{Eq. 4}$$

where b_i are some coefficients that have to be estimated.

A problem which is posed is the selection of the best hedonic model. So, in order to estimate prices a Sliced Inverse Regression (SIR) is performed, without knowing the shape of the function [14]. Then a Local Polynomial Regression (LPR) is applied and a possible shape of the hedonic function is extracted. In order to find out which is the best model, from a variety of candidate models that describe a product with a set of characteristics, the following equations are used:

$$\hat{\sigma}^2 = \frac{\{Y - \hat{f}(X\hat{b})\}' \{Y - \hat{f}(X\hat{b})\}}{n} \quad \text{Eq. 5}$$

$$d((\hat{f}, \hat{b}, \hat{\sigma}^2) = E_0\{-2\log f(Y)\} \quad \text{Eq. 6}$$

where $f(Y)$ shows the possibility for the candidate model and E_0 shows expectation under the true model. Among several candidate hedonic models the best one is derived by applying the (AIC)Akaike Information Criterion (Naik and Tsai 2001) by the following equation:

$$AIC_c = \log \hat{\sigma}^2 + \frac{1 + tr(\hat{H}_p + \hat{H}_{np} - \hat{H}_p \hat{H}_{np})/n}{1 - \{tr(\hat{H}_p + \hat{H}_{np} - \hat{H}_p \hat{H}_{np}) + 2\}/n} \quad \text{Eq. 7}$$

where $\hat{H}_p = \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}', \hat{V}$ is obtained by replacing b^* and $\tilde{f}.in\tilde{V}$ with estimators $\hat{b}and\hat{f}$, and $\hat{H}_{np}isH_{np}$evaluated at $Xb = X\hat{b}$.

Across the candidate models in several shapes of link function, the one which gives the smallest AIC is the suitable.

Finally when there are N telecommunication products in period m and $m+1$, the proposed hedonic price index can be calculated by the following equation:

$$I_{m+1/m} = \hat{f}_{m+1}(\hat{b}_i * X_i) - \hat{f}_m(\hat{b}_i * X_i) \quad \text{Eq. 8}$$

III. Price indexes for leased lines

Markets of high technology products and services, such as telecommunications, are described by fast technological changes and rapid generational substitutions. As a result, the question arisen is generally about prices of products and what influences or determines them. Particularly, how prices for products that enter the market for the first time or have been modified can be estimated, how can someone determine a price index for these products in a specific period m and what do these prices tend to become over time? For this purpose, the work presented in this section, deals with the construction of a price index for telecommunication services (leased lines) with a hedonic approach. A leased line is a permanent connection between two telecommunications sites. The prices usually depend on the distance and on the transmission rate and the operators guarantee better access to the network. The importance of such an approach is especially significant for markets characterized by rapid technological and generational changes[21].

A. Methodology

The hedonic approach is based on the fact that there is a set of consumers who have preferences over some characteristics of a service. The term 'hedonic methods' refers to a 'hedonic function' Y used in econometrics, where

$$Y = g(\beta X) + u \quad \text{Eq. 9}$$

with Y refers to data (e.g.. prices of products), X_i is a vector of regressor values (e.g. characteristics associated with the variety of the products) and u is distributed normally around zero.

In order to choose the hedonic function that associates the observed output with the vector of the variables, a number of mathematical techniques have been used, resulting in the selection of the model using an improved Akaike Information Criterion – AIC [15], by minimizing the Kullback - Leibler distance [8]. This procedure results in simultaneously choosing the relevant regressors, and a smoothing parameter for the unknown hedonic function. These techniques are extensively described in [8] and consist of the following steps:

- Firstly, the application of Sliced Inverse Regression, (SIR), in order to obtain a consistent estimate of the parameters of the model, $\hat{\beta}_{SIR}$, without requiring estimation of the hedonic function [16], [14] .
- The application of a Local Polynomial Regression, (LPR), with a Gaussian kernel, in order to estimate the unknown hedonic function by $\hat{g}(t)$, where $t = X\hat{\beta}_{SIR}$ [26].
- Finally, the application of the improved Akaike Information Criterion which minimizes the expected Kullback - Leibler distance, in order to select the appropriate model from a wider class of candidate models [8].

B. Evaluation

Evaluation of the proposed methodology was performed over **all** European countries, with data from year 1997 to year 2003, which actually includes 42 combinations of capacity, distance and

price, over time, whereas a price index is constructed for the case of Greece. These two characteristics (capacity (DIST) and distance (MB)) are widely used from telecom operators for valuating and selling leased lines across Europe. Three different distances are covered, namely 2 km (local circuits), 50 km and 200 km as well as four types of leased lines circuits, namely 64 kb/s, 2 Mb/s, 34 Mb/s and 155 Mb/s but there are enough price data only for 64 Kb/s and 2Mb/s, ensuring the compatibility of the data. All prices are presented in Euros (€) per year, excluding VAT. This overview about data prices and circuits expressed in '*Report on Telecoms Price Developments from 1997 to 2000'* which is prepared for European Commission [17].

By dividing the dataset to slices and performing the SIR algorithm, the corresponding SIR directions were calculated, after conducting eigenvalue decomposition, with respect to the covariance matrices. The plot of Y against the SIR variates is depicted in Figure 1.

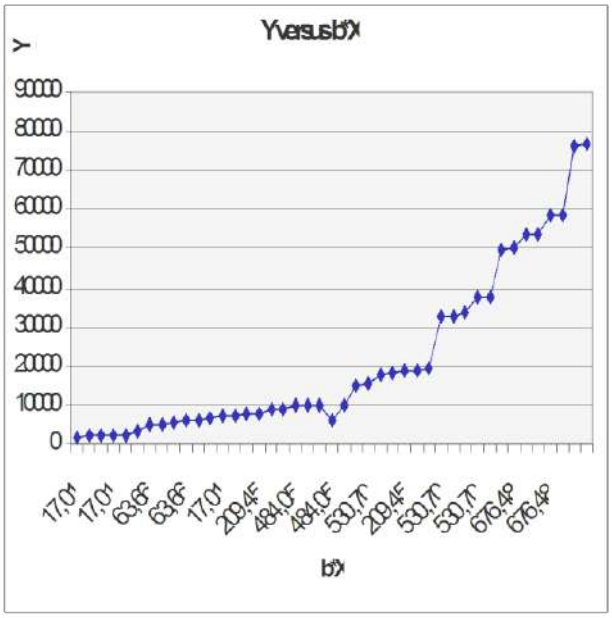


Figure 1: Plot of Y (prices) against the SIR variates

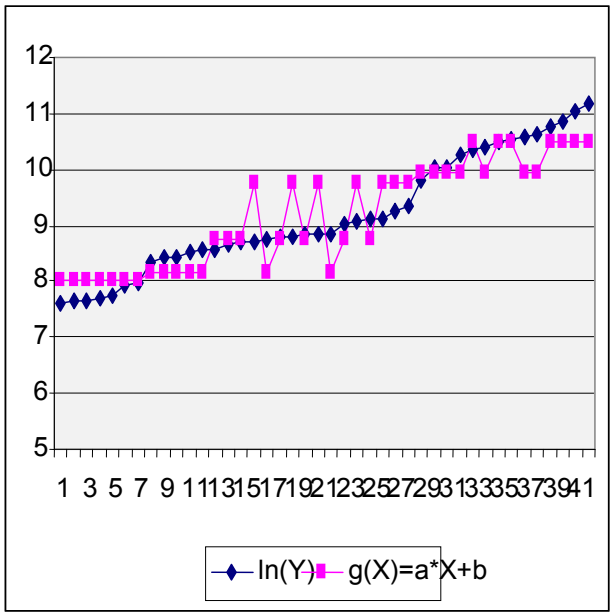


Figure 2: Selected model fitting results

This plot provides a graphical summary, useful for revealing the regression structure, thus giving an insight of the form of the underlying model. As a next step, the LPR algorithm is applied (Figure 2), using different bandwidths (i.e. 5, 10 and 100). Taking into account the shape of the above plot the following hedonic function will be evaluated by the AIC,

$$Y = \ln(P_i) = g_0(\beta_0 X_0) \qquad \text{Eq. 10}$$

where P_i is the price of a product variety and $g(\beta_0 X_0)$ could be:

i) $g(\beta_0 X_0) = \alpha + \beta_0 X_0$

ii) $g(\beta_0 X_0) = \exp(\alpha + \beta_0 X_0)$

iii) $g(\beta_0 X_0) = \alpha + \beta_0 X_0^2$

iv) $g(\beta_0 X_0) = \alpha + \beta_0 \ln(X_0)$

Concluding, the considered equation relating the product's price and its characteristics is the following:

$$\ln(P_i) = \beta_0 + \beta_1 Dist + \beta_2 MB \qquad \text{Eq. 11}$$

where β_i are the coefficients estimated in the above described procedure.

Therefore, the proposed hedonic price index can be calculated by the following equation:

$$I_{m+1/m} = \hat{g}_{m+1}(\beta_0 + \beta_1 Dist + \beta_2 MB) - \hat{g}_m(\beta_0 + \beta_1 Dist + \beta_2 MB) \quad \text{Eq. 12}$$

This index and its evolution for several years are presented for the case of the Greek market (Figure 3) for the telecommunications leased lines, showing the decrease of the prices in 1998 as a part of market liberalization and the stability of this market in the next years.

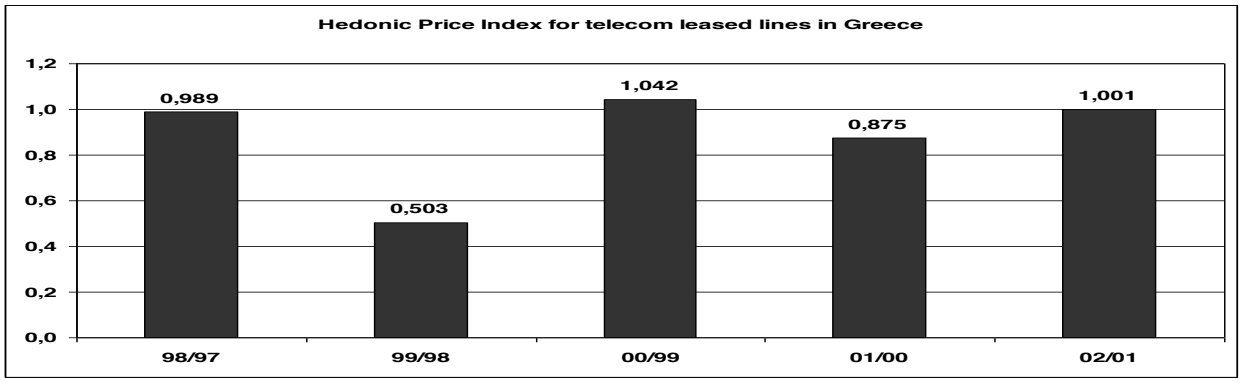


Figure 3: Hedonic price index evolution for the case of leased lines market in Greece

IV. NON PARAMETRIC APPROACH FOR ADSL CONNECTIONS

It is evidence that the demand for Asymmetric Digital Subscriber Line (ADSL) connections increases day by day in all European countries as much as worldwide and it is well known that Internet's penetration is considerably amazing. At the same time, due to the fierce competition among ADSL connections providers, several packages are offered in attractive tariffs. As a product consists of various characteristics that consumers prefer, the questions that arise are how should consumers' choices and preferences for ADSL connections affect tariffs and what are the more significant and powerful characteristics that shape tariffs of ADSL connections? This work in this section provides a hedonic price analysis of ADSL connections for the European market. The problem which is posed is the selection of the best hedonic model. So, in order to estimate prices a sliced inverse regression (SIR) is performed, without knowing the shape of the function. Then by applying Local Polynomial Regression (LPR) a possible shape of the hedonic function is given. Among several candidate hedonic models and by applying AIC criterion, the best one is derived [22].

A. ADSL High-speed Internet connections

Broadband services and applications are classified according to the offered data rate. The domination of ADSL technology for broadband access across Europe during the last years demonstrated the high-speed Internet access and IP-telephone as the most common broadband services.

In order to specify the basic basket of broadband services an extended survey of providers across European countries took place, focused on the services offered, the pricing policy as well as the development in broadband market. As a result, the typical ADSL service basket was determined as a combination of main and additional services according to Figure 4. Main services are distinguished into "horizontal" and "vertical" according to the number of fixed variables among supported data rate (DR), maximum consumed data volume (V) and maximum allowed minutes on line (T). Additional services includes a number of email addresses, web space for web hosting and/or file storage and optional free local phone calls and static IP address. The choice of the appropriate combination for each operator is depending on the specific business plan as well as the techno-economic model parameters and assumptions.

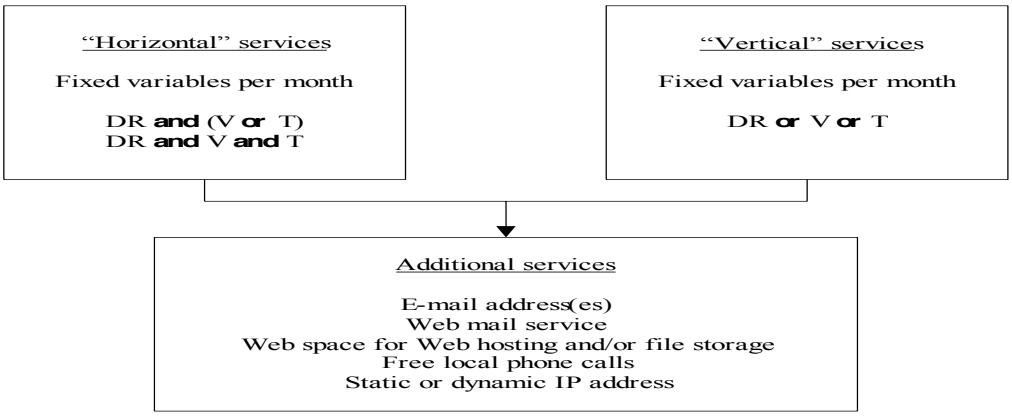


Figure 4: Typical ADSL service basket

The evolution of broadband technology offers new and challenging options. The EU Commission's "Broadband for all" policy is expected to grow the interest for broadband in the next years and to enforce the infrastructure competition among providers. As a result of this competitive environment the provision of enhanced broadband services with reduced tariffs is expected to increase significantly the number of broadband subscribers.

B. Evaluation of the model

Evaluation of the methodology was performed over 15 European countries and more specifically, tariffs' data have been collected over the period from 2003 to 2005. Apart from tariffs, data that specify ADSL connections, such as supported data rate DR (up and down), maximum consumed data volume (V) and maximum allowed minutes on line (T) have also been collected and analyzed.

Considering that the hedonic function is given by the following equation:

$$P_i = f(X_i) \qquad \text{Eq. 13}$$

with P_i is the price of a variety (or a model) i of a product and X_i is a vector of characteristics associated with the variety.

First of all, the SIR algorithm is applied, where the data are sorted by P_i and then are divided into three slices. Without specifying the unknown link function we derive:

$$\hat{b}_{SIR} = (0.581899, -0.78326, -0.00011, -0.21886)$$

The above results imply that the price is strongly related only with the downlink DR [18].

By having four characteristics we take under consideration 4^2-1 nested candidate models. For each of the nested models SIR estimates are obtained (Figure 5) and then by applying the LPR (Figure 6) the link function $\hat{f}_k\,(k = 1,2,3,..15)$. Figure 5,6 also shows that some individual tariffs are existing which decline significantly from the main cluster.

In order to examine the relationship between the price of an ADSL connection and their main characteristics, such as the supported data rate and the maximum consumed data, several candidate link functions are applied. Across the candidate models in several shapes of link function, the one which gives the smallest AICc (0,2125746 vs 7,846617) value from equation 13, is described by the equation:

$$\ln(p) = b_0 + b_1 \cdot DR(down) + b_2 \cdot DR(up) + b_3 \cdot (V) + b_4 \cdot (T) \qquad \text{Eq. 14}$$

Even link functions such as hyperbolic sine or hyperbolic cosine give almost the same results with the non linear functions. Although linear model on logged price is not comparable with all the other models, because of the AICc value, linear model on logged scale has an advantage.

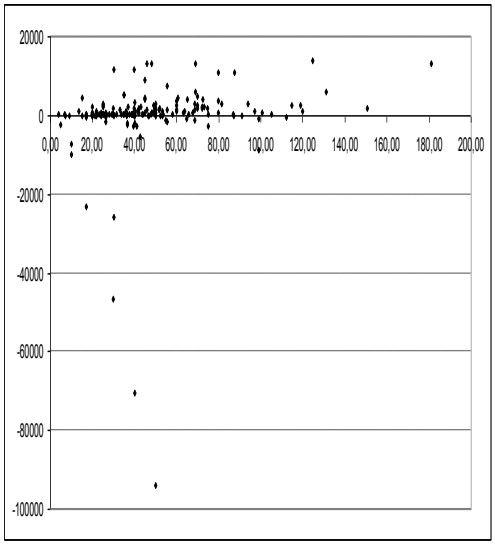


Figure 5: Plot of P_i against SIR directions

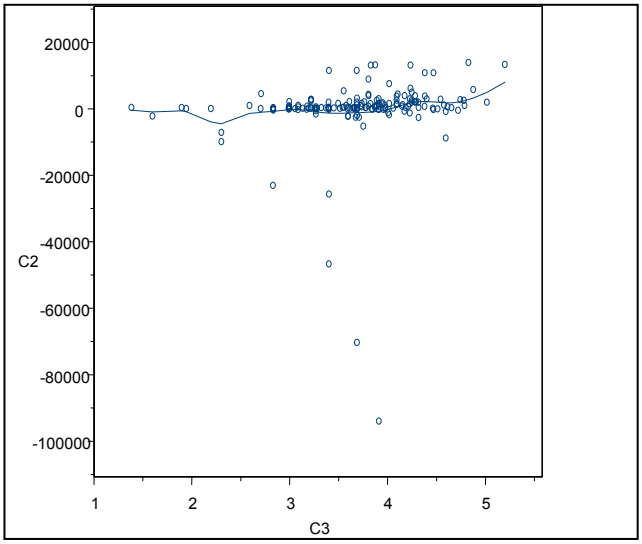


Figure 6: Local Polynomial Regression with Kernel Smoothing

Working in the logarithmic scale using a linear model results shows a better fit than all the other models because first of all the residuals from the log-linear model are all around zero (Figure 7) [18] and less standard error (0,5845 vs 26,58).

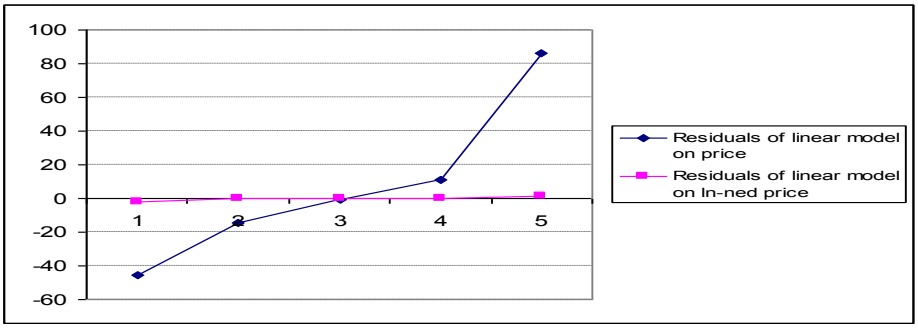


Figure 7: Residuals

Using data such from all European countries, it can been observed that it is not easy to compare prices for different data rate and consumed volume and allowed time on line, but there are similarities and patterns that must be evaluated. It is obvious that the more a consumer demands for a product with upper services, the more the prices of this product are increased. Because of this consumers' 'behavior' there is no implicit prices for all characteristics.

V. INCLUDING SOCIO-ECONOMIC VARIABLES IN HEDONIC MODEL

During the last decade broadband services and applications achieved significant penetration into the mass market across Europe. As the operators will continue to improve network infrastructures, customers are expected to "enjoy" new services in more attractive prices. Telecommunications' pricing process is influenced by a number of factors, such as subscribers' profile, market competition as much as users' income. In this work an overview of tariffs for ADSL connections across Europe is presented and a hedonic model is applied in order to identify and estimate the characteristics which are of substantial influence of the prices' shape. Finally, a price index for ADSL connections is constructed [23].

A. Socio-economic variables

According to the economic theory described above regarding hedonic models, there are several characteristics that influence tariffs. Data by the meaning of attributes of a given product can be divided in three categories: spatial, physical and socio-economic [19].

The spatial division can be handled separate by location. Attributes or exogenous variables are made up of physical quality characteristics and socio-economic attributes. Measures and differences that compare geographical areas have implications for regional cost of living. This has a substantial effect not only to consumers' choices and preferences, but also to the governments' policies and market relocation as well. This is an index for decision-making affecting all involved parties.

The physical quality characteristics are the attributes which are unique and specify the product. For instance, a bundle of physical characteristics can be: the offered data rate, the maximum consumed data volume etc.

The socio-economic variables are characteristics that are not easily calculated, but it is assumed that influence the ADSL market expansion and penetration. Such characteristics are: the educational level, the age, the income and the personalization of the subscribers, the technological infrastructure improvement, the year of telecommunications' liberalization etc. As to the customers' interests, they can be identified from the web pages they visit and the amount of time they spend on them. But it is extremely difficult to gather data, in order to understand customers' profile and characteristics, because on one hand the veracity of them is doubtful and on the other customers are not interested in providing information for privacy concerns. By including in the model the socio-economic variables implicit market's prices may arise for quality characteristics.

B. Evaluation of the model

Following the theoretical presentation of the preceding sections, a hedonic (econometric) model was applied, in order to study a typical ADSL basket across the European countries. The evaluation was based on data for ADSL connections, which were collected from year 2003 to 2005 for both residential and business connections. Of course there is a large variation of values of the participating variables, as well as in price levels between the participating countries, for the same service considered. Evaluation data were collected from Austria, Belgium, Denmark, Finland, France, Germany, Iceland, Italy, Netherlands, Norway, Spain, Sweden, Switzerland, UK, and Greece. In addition, they correspond to all operators, no matter the number of subscribers they have. By applying a multiple regression model to the ADSL connections data, the best fitting function for every year (2003, 2004 and 2005), turned out to be:

$$P_{im} = b_0 + b_1 \cdot D(\text{Mbps})_i + b_2 \cdot U(\text{Mbps})_i + b_3 \cdot \ln V(\text{Gbps})_i + b_4 \cdot \ln(GDP(€)^2)_i + b_5 \cdot \ln((OP)^2)_i + b_6 \cdot DUR_i \quad \text{Eq. 15}$$

where D and U is the downstream and upstream data rate in Mbps, respectively, V is the maximum consumed data volume in Gbps, GDP is the Gross Domestic Product of each country in €, OP is the number of operators in each country and DUR is the downstream to upstream ratio and m is the year. Finally, coefficients b_0 to b_6 are constants and they are the corresponding weighting factors. Table 1 shows the results of fitting by a multiple linear regression model, in order to describe the relationship between the subscription price and the six aforementioned identified variables. The standard error of the estimation shows the standard deviation of the residuals, where the Mean Absolute Error (MAE) is their average value. It is noticeable that the highest P-value of the independent variables is 0.2258 belonging to $\ln(GDP^2)$ [20].

The P-value that appears in the results' table is related to the probability for the corresponding parameter to be equal to zero. So, this parameter can be considered as less significant for the model, the evolution of broadband tariffs in the presented case. Finally, since the coefficient determination (R-squared) is in excess of 94% the validity of the above model is even more strengthened.

Using data such, for the case of Europe, it can be observed that it is not easy to compare prices for

different characteristics such as data rate, consumed volume, allowed time on line, since on the same time, the subscription price for year 2004 is 50% less than this of 2003, but for the period 2004 – 2005 it is approximately the same.

Table 1: Regression analysis

REGRESSION ANALYSIS

Dependent variable: Subscription Price (€)			
Parameter	Estimated Coefficients	Standard Error	P-Value
Downstream (Mbps)	-26.01	8.38	0.0210
Upstream (Mbps)	300.30	58.54	0.0022
ln(Operators^2)	9.26	3.68	0.0456
ln(GDP^2) (€)	-15.16	11.23	0.2258
Downstream to Upstream ratio	6.178	2.34	0.0388
ln(Volume) (Gbps)	15.28	4.46	0.0141
R-squared: 94.84 %			
Standard Error of Estimation: 7.55			
Mean absolute error: 3.49			

However, in countries such as Belgium and The Netherlands, prices were reduced during 2003 – 2004 but in 2004 – 2005 remained almost the same, whereas in other such as Spain there is a continuous reduction across years.

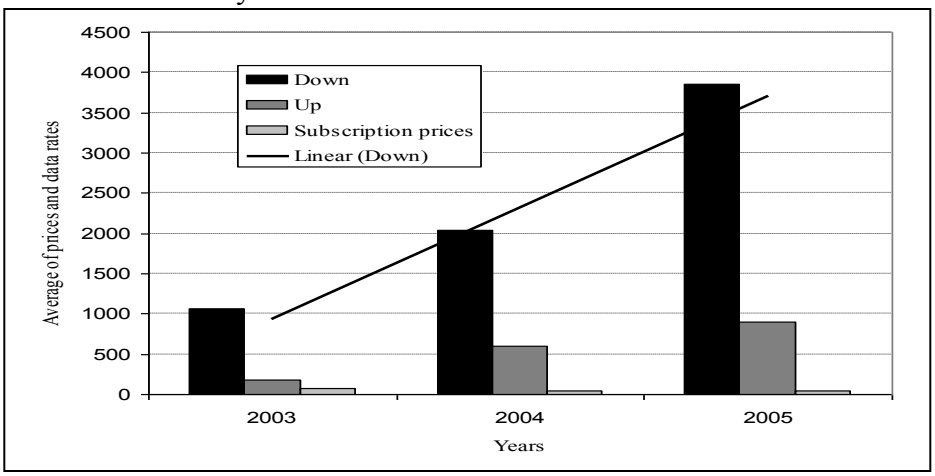


Figure 8: Evolution data rate (down) in Europe

Therefore, the average value of prices of ADSL connections and their characteristics (downstream and upstream) was computed for all countries (Figure 8), in order to provide an estimation for prices and physical characteristics (downstream and upstream) of a typical European ADSL connection. Once more, it can be observed that downstream and upstream data rates seem to follow a linear path over years (solid line in Figure 8).

Finally, in the case that there are N telecommunication products in the period between m and $m+1$, by considering Equation 4 which describes the relationship between prices and characteristics, in a variety of ADSL products, the proposed hedonic price index is calculated by Equation 8.

This behavior fits to the hedonic approach and it can be observed by calculating the hedonic price index from equation 16. In addition, In Figure 9, the calculated index and its evolution are presented for the case of European countries broadband market. It is observed that for the period 2003 – 2004 there is a trend of 50% reduction in prices.

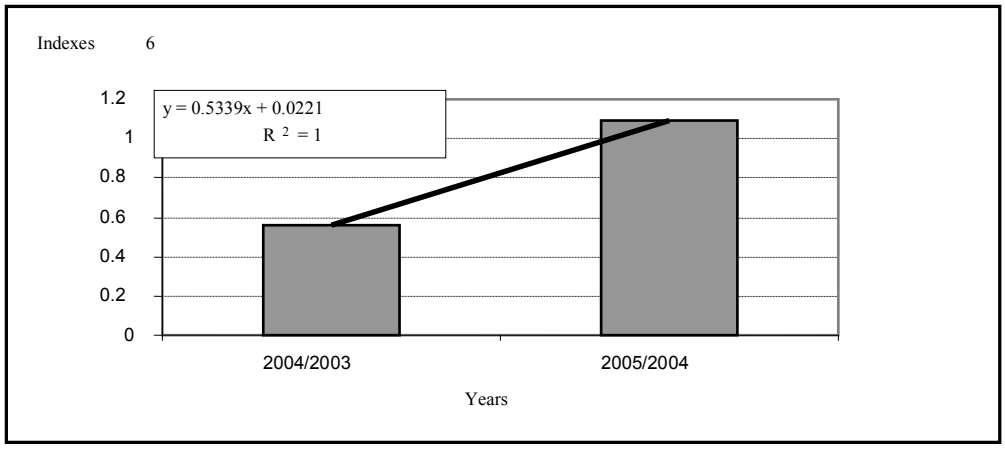


Figure 9: Hedonic price index evolution for the case of ADSL market in Europe

However, this does not apply for the period 2004 – 2005, even if the price index has approximately the value 1, which means that prices for ADSL services are the same. Taking into account that there is a significant trend in downstream and upstream bandwidth, it can be extracted that there is an implicit reduction of prices, since operators offer better services at more or less the same price.

VI. CONCLUSIONS

The main conclusions drawn by this doctoral thesis, which also constitute its contribution, lie in the following points:

- Overview of the physical, social, economic and spatial parameters that influence the prices of products-services and networks of the telecommunications market.
- Development of a methodology for the definition of characteristics, of telecoms products-services and networks that affect their prices.
- Categorization and grouping of characteristics of access technologies.
- Quantification of telecoms' networks and products data related to characteristics and the emergence of influence of them on the telecoms products-services pricing.
- Construction of price indices and clarification of the trend of prices and indices of the telecoms products and services.
- Implementation of methodological approach on different telecoms products-services and networks.
- Integration of price indices into the demand models and assessment of their influence on the diffusion process of telecoms products and services.

VII. REFERENCES

[1] D.W.Cartwright & S.D. Smith (1988), "Deflators for Purchases of Computers in GNP: Revised and Extended Estimates 1983-88", *Survey of Current Business, November*

[2] Z.Griliches (1961), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change" in *The Price of the Federal Government*, General Series no. 73 , Columbia University Press for NBER, New York , 137-196.

[3] N.Jonker (2001), "Constructing quality adjusted price indexes: a comparison of hedonic and discrete choice methods", *De Nederlandsche bank, Econometric Research and Special Studies Department..*

[4] R.Cole (1986) "Quality –adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment", *Survey of Current Business*, Bureau of Economic Analysis, US Department of Commerce.

[5] J.E.Triplett & R.J.McDonald (1977), "Assessing the Quality Error in Output Measures: The Case of Refrigerators", *Review of Income and Wealth*, 137-176.

[6] A.Moreau. (1991), "A price Index for Microcomputers in France' , Document de *Travail de la Direction des Statistiques Economiques*, 9109..

[7] J.E.Triplett (2000), "Draft copy Handbook on quality adjustment of price indexes for information and communication technology products", *OECD, Paris*.

[8] P.A.Naik & C.L.Tsai (2001), "Single –index model", *Biometrica Trust ,88,3,pp. 821-832*.

[9] F.V.Waugh (1928**)**, "Quality Factors Influencing Vegetable Prices", *Journal of Farm Economics*, 10(2), 185-196.

[10] A.T.Court (1939), "Hedonic Price Indexes with Automotive Examples", *The Dynamics of Automobile Demand,* The General Motors Corporation, New York, 99-117.

[11] R.Stone (1956), "Quantity and Price Indexes in National Accounts" *Organization for European Economic Cooperation*.

[12] R.Stone (1954), "The Measurement of Consumer Behaviour and Expenditure in the United Kingdom, 1920-1938*", Studies in the National Income and Expenditure of the United Kingdom.*

[13] E.R.Berndt (1991), "The practice of econometrics: classic and contemporary"*. (Addison-Wesley Publishing)*.

[14] K.C.Li (1991), "Sliced inverse regression for dimension reduction (with Discussion)", *J. Am. Statist. Assoc. 86, 316-42.*

[15] H.Akaike (1973). "Information theory and an extension of the maximum likelihood principle", *2nd Int. Symp. Inform. Theory, suppl. Problems of Control and Inform. Theory,* 267-281.

16] N.Duan and K.C.Li (1991), "Slicing regression: a link free regression method". *Annals of Statistics* 19, 505-530.

[17] Teligen (2000), "Report on Telecoms Price Developments from 1997 to 2000" European Commission.

[18] J.E.Triplett (2004), "Hand book on hedonic indexes and quality adjustments in price indexes', *OECD science, technology and industry working papers, 2004/9*.

[19] J.Chowhan, M. Prud'homme, (2004) "City comparisons of shelter costs in Canada: a hedonic approach", *Prices Division*.

[20] S.Goodman (1999), "Toward evidence-based medical statistics. 1: The P value fallacy", *Annals of Internal Medicine*, 130 (12), 995 – 1004

[21] D.Varoutas, C.Deligiorgi, Ch.Michalakelis and Th.Sphicopoulos, "A hedonic approach to estimate price evolution of telecommunication services: Evidence from Greece", Applied Economic Letters, vol. 15, no 14, pp 1131 – 1134, Nov 2008.

[22] C. Deligiorgi, C. Michalakelis, A. Vavoulas, and D. Varoutas, "Nonparametric estimation of a hedonic price index for ADSL connections in the European market using the Akaike Information Criterion," *Telecommunication Systems*, vol. 36, pp. 173-179, Dec 2007.

[23] C.Deligiorgi, A.Vavoulas, C.Michalakelis, D.Varoutas, and T.Sphicopoulos, "On the construction of price index and the definition of factors affecting tariffs of ADSL connections across Europe," Netnomics DOI:10.1007/s11066-008-9023-0, vol. 8, pp. 171-183, 2008.

[24] S.Rosen (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy*, (92), 34-55.

[25] M.Okamoto & T.Sato (2001), "Comparison of hedonic method and matched models method using scanner data: the case of PCs, TVs and digital cameras" *Sixth Meeting of International Working Group on Price Indices, Canberra, Australia.*.

[26] J.S.Simonoff (1996), 'Smoothing Methods in Statistics' *New York: Springer*

.

Performance Analysis of Cognitive Radio Networks Using Cross-layer Design Approaches

Fotis T. Foukalas*

National Kapodistrian University of Athens
Department of Informatics and Telecommunications
foukalas@di.uoa.gr

Abstract. This thesis studies the performance of both opportunistic spectrum access (OSA) and spectrum sharing (SS) cognitive radio networks (CRNs) using cross-layer design (CLD) approaches in order to provide reliable and optimum packet transmission at the medium access control (MAC) layer with quality of service (QoS) guarantees, to optimize the secondary user (SU) performance given the primary user (PU) protection and to realize the impact of imperfect spectrum sensing at the MAC layer. The reliable and optimum packet transmission at the MAC layer with QoS guarantees is accomplished by the combination of adaptive modulation with hybrid automatic repeat request (HARQ) protocol in OSA CRNs. The SU's performance is maximized by formulating a convex optimization problem for the SU's capacity over the power control (PoC) and spectrum sensing (SpSe) that is solved with an iterative subgradient method which results in optimal power allocation and sensing threshold selection. Finally, the impact of imperfect SpSe at the MAC layer is realized by modelling the Carrier Sensing Multiple Access with Collision Avoidance (CSMA/CA) and the SpSe as two state Markov processes and thus obtaining the joint steady state distribution that encompasses the parameters of both mechanisms.

1 Introduction

Third-generation (3G) and beyond 3G mobile communication systems must provide interoperability with the Internet, increase throughput for mobile devices, and optimize their operation for multimedia applications. The limited ability of traditional layered architectures to exploit the unique nature of wireless communication has fostered the introduction of cross-layer design (CLD) solutions that allow optimized operation for mobile devices in the modern heterogeneous wireless environment. In this thesis, we first present the major cross-layer design solutions that handle such problems, and discuss cross-layer implementations with a focus on functional entities that support cross-layer processes and the respective signaling. In addition, we consider the associated architectural complexity and communication overhead they introduce. Furthermore, we point out the major open technical challenges in the cross-layer design research area. Finally, we conclude our article with a summary of cross-layer approaches developed thus far and provide directions for future work [1].
Besides, cognitive radio is considered as one of the most important enablers for achieving enhanced spectral efficiency in wireless communications. In the sequel, in this thesis, we present a cross-layer design for reliable data transmission over a cognitive radio network which combines adaptive modulation at the physical layer and hybrid automatic repeat request at the data link layer. The cognitive radio network follows the principles of opportunistic spectrum access that utilises an optimal power adaptation policy for channel allocation. The obtained numerical results denote that the considered approach achieves significant spectral efficiency improvement and therefore it could be deployed in wireless communication networks that encompass cognitive capabilities. Furthermore, we assess the introduced interference and we show that it can be kept within levels that do not jeopardise our design [2].

* Dissertation Advisor: Lazaros Merakos, Professor

Furthermore, we study the problem of maximizing spectral efficiency of cognitive radio network deployments subject to an interference constraint and under specific quality of service (QoS) guarantees. The interference constraint corresponds to the upper limit of the received power that can be tolerated at the licensed users' due to transmissions from unlicensed users. The QoS guarantees stem from the requirements imposed by the applications running at the users' terminals. A cross-layer design is adopted that maps the user's requirements into delay related QoS guarantees at the data link layer and error probability QoS guarantees at the physical layer. The obtained numerical results provide important insights regarding the impact of the considered constraint and guarantees on the achievable spectral efficiency of cognitive radio networks [3].
Moreover, a joint optimal power allocation and sensing threshold selection for capacity maximization at the secondary user (SU) in spectrum sharing (SS) cognitive radio networks (CRNs) is proposed. Hence, both optimal power allocation and spectrum sensing is considered in the SS CRNs model. The obtained results show that such a joint optimal selection improves the performance of the SU by maximizing its capacity [4]. Besides, we propose capacity optimization through sensing threshold adaptation for sensing-based cognitive radio networks. The objective function of the proposed optimization is the maximization of the capacity at the secondary user subject to transmit power and sensing threshold constraints for protecting the primary user. After proving the concavity of capacity on sensing threshold, the problem is solved using the Lagrange duality decomposition method in conjunction with a subgradient iterative algorithm. The numerical results show that the proposed optimization can lead to significant capacity maximization for the secondary user as long as this is affordable to the primary user [5].
Finally, we introduce a cross-layer design (CLD) of carrier sensing multiple access with collision avoidance (CSMA/CA) at the medium access control (MAC) layer with spectrum sensing (SpSe) at the physical layer for cognitive radio networks (CRNs). The proposed CLD relies on a Markov chain model with a state pair containing both the SpSe and the CSMA/CA from which we derive the transmission and collision probabilities. The derived probabilities can be used as performance criteria to evaluate the performance of specific CRNs when they are deployed in a distributed coordination fashion that is prone to collisions [6].
The rest of this document is organized as follows. In Section 2, we describe the major results of this thesis that are namely the spectral efficiency of CRNs under interference constraint and with QoS guarantees, the joint optimal power allocation and spectrum sensing threshold selection for SS CRNs and the CLD of CSMA/CA with SpSe in CRNs. We conclude this document with the overall conclusions derived from the investigations accomplished within the framework of this thesis.

2 Performance analysis of cognitive radio networks using cross-layer design techniques

In this section we present two main results of this thesis. First, the spectral efficiency of CRNs under interference constraint and QoS guarantees and second the joint optimal power allocation and sensing threshold selection for SS CRNs

2.1 Spectral efficiency of cognitive radio networks under interference constraint and QoS guarantees

In this section we focus on the approach in order the SU be able to share efficiently the same spectrum band with PU via CR and we model the respective procedures. To this

end, this study reveals the impact of the adoption of an interference constraint at the PU on the spectral efficiency of the SU.

We assume a CR networking layout with a PU and a SU as depicted in Fig. 1 where the channel allocation technique that is adopted follows the principles of the spectrum underlay paradigm of CRNs since it encompasses an interference power as we will describe below. Both transmitters and receivers are considered as users with the notation PU-Tx and PU-Rx for users of the primary network (PN) and SU-Tx and SU-Rx for the users of the secondary network. The channel power gains from SU-Tx to SU-Rx and PU-Rx are denoted by γ_{11} and γ_{12}, respectively, and from PU-Tx to PU-Rx and SU-Rx by γ_{22} and γ_{21}, respectively. Channel power gains γ_{12} and γ_{21} are considered as interference for PU-Rx and SU-Rx respectively [7].

With the considered channel allocation technique both PU-Rx and SU-Rx are using the same frequency band (e.g. f_1) and henceforth both should transmit in a way that achieves a reasonably high transmission rate without causing too much interference to each other [7]. Besides, we consider that the SU-Rx exploits spectrum sensing to discover more than one spectrum bands in the wideband range $\{f_1, f_2, ..., f_8\}$ and thus it is able to settle in more than one spectrum bands [8]. Spectrum sensing is one of the most important tasks for terminals with cognitive capabilities and that is why the corresponding users are denoted as cognitive ones [9]. It should be noted that the considered channel allocation technique give access priority to PU-Rx as implemented in a spectrum sharing system in general.

For studying the maximization of the spectral efficiency of SU-Rx that is subject to the transmit and interference power constraints, we analyze first the channel allocation technique. Both constraints are related to SU-Tx power transmission as we will discuss later. Then, we continue with the technique, by which the SU-Rx is able to sense the available spectrum bands offered in a wideband regime, in order to present its spectral efficiency per unit bandwidth. This means that the network can simultaneously offer more sub channels and thus a gain is manifested in terms of spectral efficiency per unit bandwidth in the whole networking system.

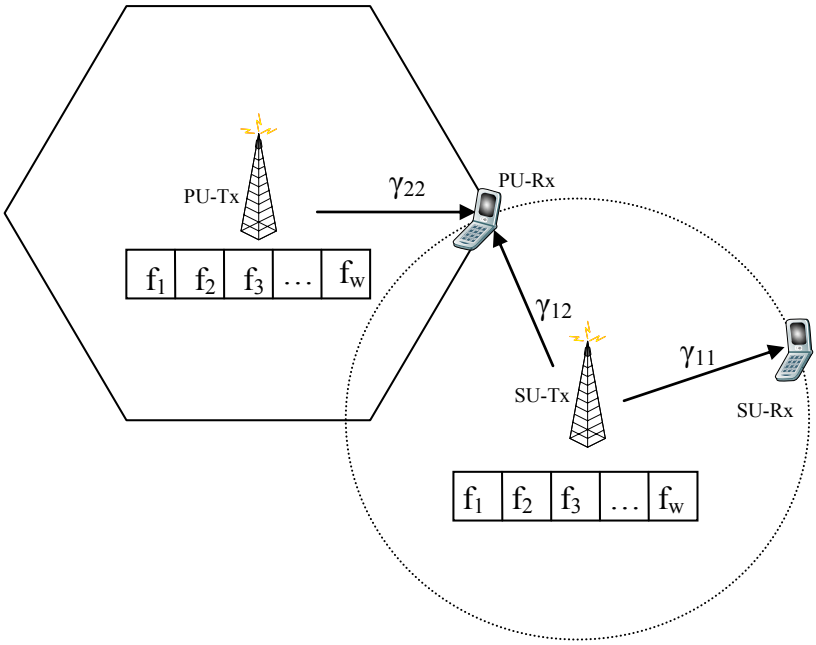


Fig. 1. Cognitive Radio Network Model

In figure 2 we present the potential gain in spectral efficiency per unit bandwidth that is achieved when the considered CRN deployment exploits also spectrum holes in the

spectrum range. As previously mentioned the SU-Rx is able to sense the spectrum bands in order to transmit throughout the sensed vacant bands. We include the aforementioned power constraints and QoS guarantees in such a CR system and we depict the spectral efficiency gain achieved. The results obtained by applying the optimal power allocation policy for the following interference power constraints levels $-10dB$, $0dB$, $10dB$ and $20dB$ are depicted with blue lines. With black lines are illustrated the results with QoS guarantees equal to $p_{loss} = 0.001$ and $R_{\max} = 0$, while the red lines show the results when $p_{loss} = 0.001$ and $R_{\max} = 4$. From the presented curves we conclude that as the received CNR increases, the network behavior tends toward the no cognition case since no spectral gain is exhibited something that was evident in figure 3 either. This can be explained from the fact that when the cutoff level γ_{11}^{*} is very low, the possibility for a vacant band is bigger than when this value is higher where more power is poured within each channel and the user with the higher priority occupies the whole bandwidth. This is accomplished (i.e. $\gamma_{11}^{*} \to 0$) indeed at lower average CNR regions when the worst case scenario is considered e.g. in case of a tight interference power constraint $-10dB$.

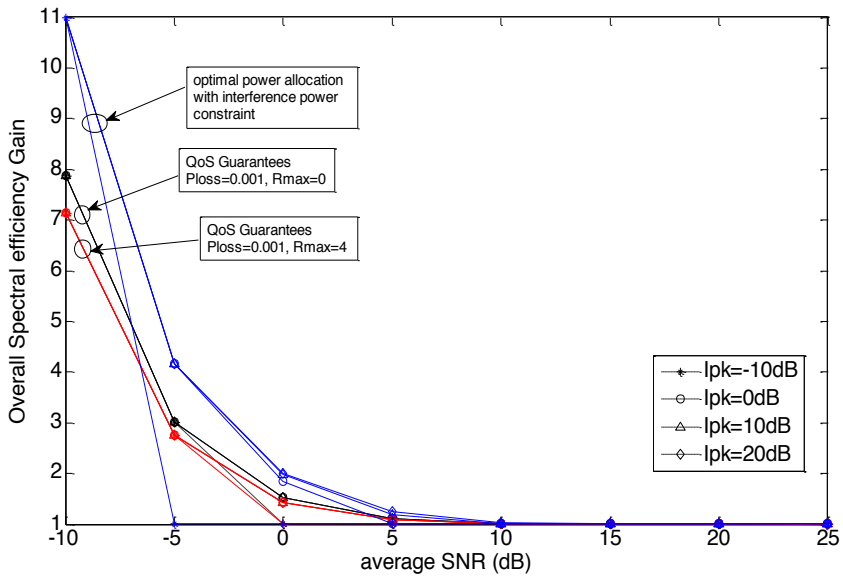


Fig. 2. Performance Gain in Spectral Efficiency under interference constraints

In general an interference power constraint affects the system in terms of vacant band i.e. band in which the threshold is above the cutoff value and it makes the system to downgrade its performance. In addition to that, the incorporation of QoS guarantees downgrade the performance gain achieved from such a system and makes it to operate as a conventional one in lower CNR regions. However, this is expected since on the other hand the constraints guarantee the QoS that can be offered by the network in order to satisfy the users' requirements. The presented approach could be extended further by assuming a multi-carrier modulation technique such as orthogonal frequency division multiple access (OFDM) that is most frequently considered for the implementation of CRNs. This extension requires a proper formulation of the analysis and assessment presented in this paper that will introduce some extra parameters (e.g. guard interval, number of OFDM symbols, etc) in order to correctly obtain the spectral efficiency gain that will now achieved per subcarrier. This extension can be considered as a future work.

2.2 Joint Optimal Power Allocation and Sensing Threshold Selection for SS CRNs

In spectrum sharing (SS) cognitive radio networks (CRNs), optimal power allocation (OPA) and spectrum sensing (SpSe) are used for the protection of the primary user (PU) from harmful interference caused by the secondary user (SU). Furthermore, for capacity maximization of the SU, the main parameters related to OPA, i.e. the SU's transmit power, P_t, is adapted according to the received signal-to-noise ratio (SNR), γ_s, and related to SpSe, i.e. sensing threshold, η, and sensing time, τ, for a given sensed SNR, γ, need to be also carefully selected [10]. Previous studies on SU's capacity maximization include SS CRNs models with SpSe [12], [13] or without SpSe [11], [10]. For the former and more general case, the optimization presented in [6] is considered over P_t and τ assuming η to be constant. In [13] although the effects of η as a variable are studied, the research is focused on the interference caused to the PU rather than the optimization of the SU's capacity. Thus, a more general approach is presented in this letter where a jointly OPA and SpSe threshold selection is considered so that the SU's capacity is maximized over P_t and η.

Figure 3 illustrates the performance of C_s obtained from the joint optimization problem in (2), versus η, for different values of γ and P_{av}. As in [7] and [13], for the performance evaluation results we have assumed that for Rayleigh fading channels the channel power gains (exponentially distributed) are assumed with unit mean, AWGN with variance $N_0 = 1$ and $\pi_1 = 0.4$. Furthermore, for the OPA, the constraint on peak interference power is assumed to be $I_{pk} = 0dB$ while for the SpSe, $\tau = 1ms$ [14]. The performance evaluation results obtained clearly show that C_s increases as γ decreases and/or P_{av} increases while its improvement becomes negligible when $P_{av} < I_{pk}$.

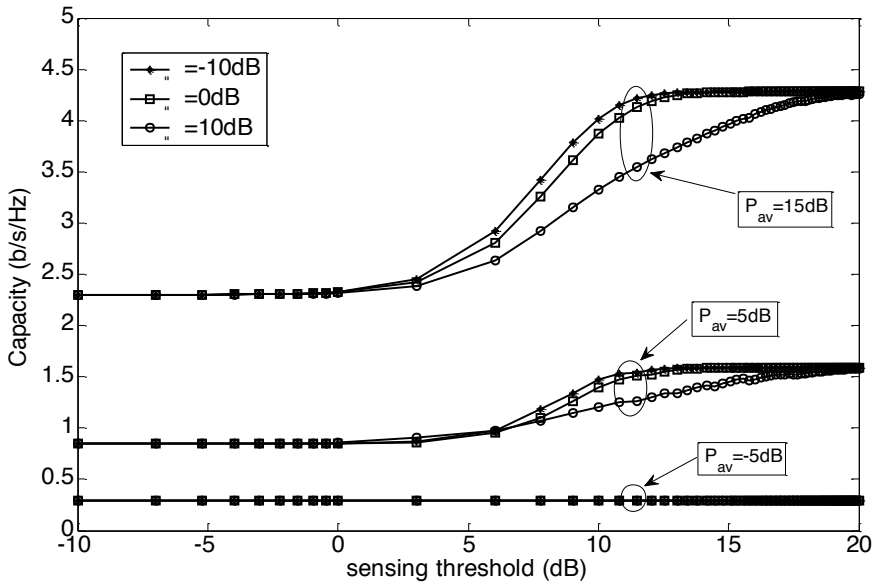


Fig. 3 Capacity C_s vs. sensing threshold η for different γ and P_{av} assuming $I_{pk} = 0dB$

Figure 4 illustrates the throughput computed as $\xi_s = (T - \tau / T)C_s$ based on the SpSe and frame transmission models that have been proposed in [6], where T is the frame duration, $T - \tau$ is the frame duration for data transmission and C_s is taken from (1). Therefore ξ_s represents the transmitted bits per frame, is the performance metric at the secondary link and it is maximized over the sensing time τ for different optimal probabilities of detection p_d^* using the maximization in (2). The performance results have been obtained for $T = 100ms$, $P_{av} = 15dB$, $I_{pk} = 0dB$ and $\gamma = -10db$. Furthermore, different target values of p_d^* are assumed that correspond to specific SpSe thresholds, η^*. However, these η^* values are identical for each target value p_d^* as depicted in Fig. 2. This is reasonable since the maximization problem is assumed over τ and not over η. This also shows that a proper selection of η for the SpSe and P_t for the OPA provides an additional C_s maximization to the one achieved by the joint optimization over P_t and τ.

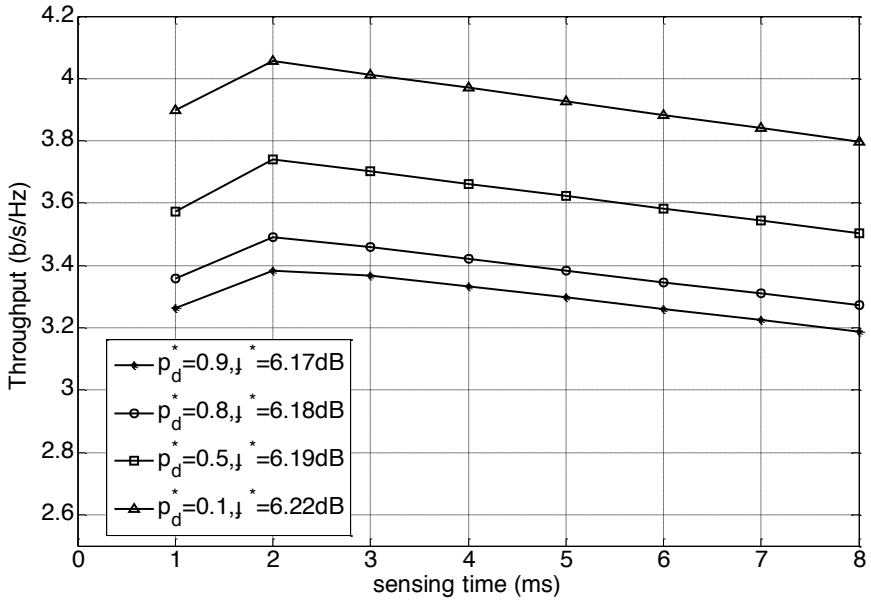


Fig. 4 Throughput ξ_s vs. sensing time, τ for different optimal p_d^* with $P_{av} = 15dB$ and $I_{pk} = 0dB$

A joint OPA and SpSe threshold selection for capacity maximization at the SU in SS CRNs has been proposed. This joint optimization leads to further capacity maximization as compared to the one achieved by joint optimization over the transmit power and sensing time. This maximization has also shown that the capacity can be further improved by properly selecting the SpSe threshold based on the sensed SNR.

2.3 Cross-layer Design of CSMA/CA with Spectrum Sensing for CRNs

In cognitive radio networks (CRNs), the channels' availability is manifested via spectrum sensing (SpSe) at the physical layer [15] and subsequently the packet transmission is accomplished through an appropriate medium access control (MAC) layer technique [16]. It has been recognized that the imperfect SpSe at the physical layer have an impact on the performance at the MAC layer [20]. To this end, the authors in [21] propose a cross-layer design (CLD) between the SpSe at the physical layer and the MAC layer in general in which a constraint on collision probability dictates the operating characteristics of SpSe. In this thesis, we introduce a cross-layer design (CLD) of SpSe at the physical layer with a specific MAC layer technique, the well-known carrier sensing multiple access with collision avoidance (CSMA/CA) protocol. Based on the CLD presented in [17], the Markov chain model of CSMA/CA presented in [18] and the Markov chain model of SpSe presented in [19], we derive the transmission and collision probabilities that can be used to evaluate the deployment of CSMA/CA protocol in CRNs. In [20] the authors present a cross-layer performance analysis of CSMA/CA in case of imperfect SpSe but they neither consider the well established exponential backoff nor a Markov chain model for multi-channel SpSe of CRNs.

We have derived both numerical and simulation results in order to validate the proposed CLD. We assume that the sensed SNR is equal to $\gamma_p = -15dB$ and the sensing time equal to $T_s = 2ms$ for primary channels with frequency $f_p = 6MHz$. Fig.5 shows the transmission probability τ (left part) and the collision probability p_c (right part) versus the number of stations n for different probabilities of detection P_d, backoff stages m and minimum contention window W. The results obtained considering one channel i.e. $C = 1$ with an activity probability equal to $a = 0.5$. The solid lines depict the case of $m = 3$ and $W = 32$, the dashed lines depict the case of $m = 3$ and $W = 64$, and the dotted dashed lines the case of $m = 5$ and $W = 32$. Since the depiction of all simulation results would make them indistinguishable, we depict them for the cases of $P_d = 1$ and $P_d = 0.9$ with $m = 3$ and $W = 64$ with circles and stars and without line. From the figure is obvious that a low probability of detection e.g. $P_d = 0.1$, results in high transmission probability τ. Furthermore, high contention window values e.g. $W = 64$ (dashed lines) and/or backoff stage e.g. $m = 5$ (dotted dashed lines) result in a lower transmission probability τ, although a high contention window value gets the transmission probability lower. We also notice that the collision probability p_c is proportional to the transmission probability τ.

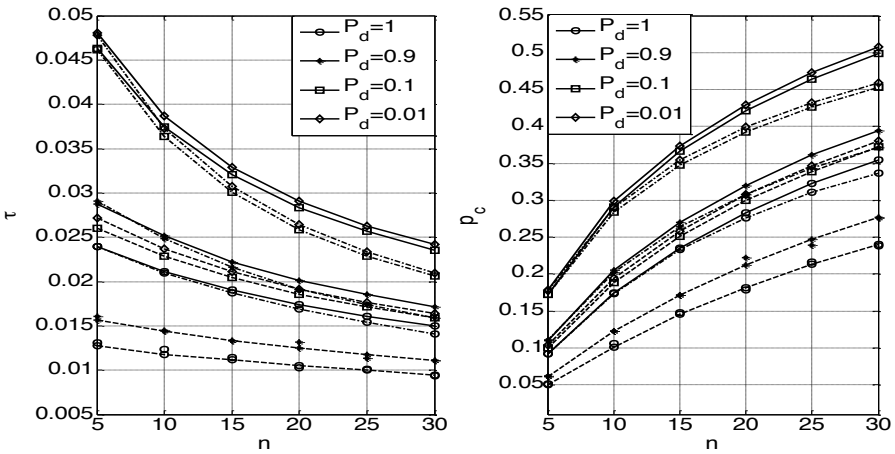


Fig. 5. Transmission probability τ (left part) and collision probability p_c (right part) vs. number of stations n for different probabilities of detection P_d with backoff stage $m = 3$ and contention window $W = 32$ (solid lines), with $m = 3$ and $W = 64$ (dashed lines) and with $m = 5$ and $W = 32$ (dotted dashed lines).

Fig. 6 shows the transmission probability τ and collision probability p_c versus the number of stations n for different values of channels' activity a and number of channels C. The results obtained considering a probability of detection equal to $P_d = 0.5$, a backoff stage equal to $m = 3$ and a contention window equal to $W = 32$. The solid lines depict the case of $a = 0$,$a = 0.5$, $a = 0.8$ and $C = 1$, the dashed lines depict the case of $C = 3$ and the dotted dashed lines the case of $C = 6$ considering the same activities for all cases. We also depict the simulation results in case of activity $a = 0.8$ and channels $C = 3$ with squares without line. Obviously, a high probability of activity e.g. $a = 0.8$ results in a lower transmission probability τ. Furthermore, for a high number of sensed channels e.g. $C = 6$, the transmission probability τ increases. As previously, the collision probability p_c is proportional to the transmission probability τ.

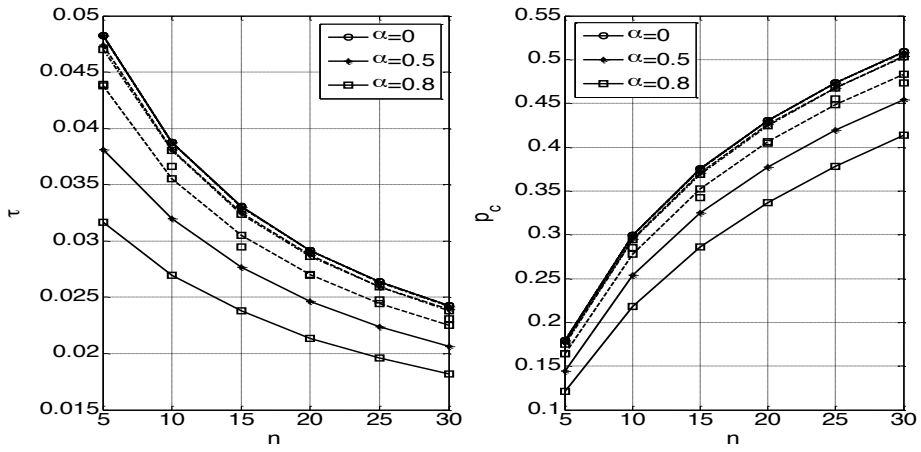


Fig. 6. Transmission probability τ (left part) and collision probability p_c (right part) vs. number of stations n for different activities a with $C = 1$ channels (solid lines), with $C = 3$ channels (dashed lines) and with $C = 6$ channels (dotted dashed lines).

3 Conclusions

In this thesis, we present a CLD over an OSA-based CRN that provides reliable data transmission. In particular, we study the achieved spectral efficiency over fading channels and the potential gain in OSA-based CRNs. The CRN model which allows for OSA to SUs by sensing the spectrum's vacant buds from the licensed network is provided. In order to provide reliable data transmission in the new wireless communication network we adopt adaptive modulation and automatic repeat request at the physical and data link layer respectively. The obtained results show the improvement in spectral efficiency versus the average SNR of the fading channel.

Finally, the introduced interference in such a network is assessed and we show that it can be kept within levels that do not affect the operation of the considered overlay communication system.
Besides, we analytically derived the maximum spectral efficiency achieved at the physical layer of a CRN under interference power constraint with QoS guarantees. Based on an optimal power allocation strategy at the physical layer of the SU we introduced delay related QoS guarantees through a cross-layer model. The specific cross-layer model combines the adaptive modulation and ARQ at the physical and data link layer, respectively that gives us the ability to evaluate the achievable spectral efficiency over fading channels. The impact of the imposed constraints on the spectral efficiency is analyzed via numerical results assuming a Rayleigh fading channel. It is shown that the spectral efficiency cannot be increased up to an average CNR value that is equal to the interference power constraint. Besides, we evaluate the gain in spectral efficiency that is defined in the case that the SU is able to sense a spectrum range of vacant channels. Finally we showed that a proper selection of the constraints and guarantees at the SU can make the operation of the overall CRN more efficient.
Furthermore, we study the capacity optimization for sensing-based cognitive radio networks over sensing threshold. In particular, we consider a sensing-based spectrum sharing CRN in which both power control and spectrum sensing are employed for the PU's protection. The proposed optimization is proved to be a convex optimization problem that we solve using the Lagrange dual decomposition method. A subgradient iterative algorithm provides the optimum values for both transmit power and sensing threshold of the power control and spectrum sensing, respectively. The numerical results show the SU's capacity maximization achieved through sensing threshold adaptation and the corresponding capacity loss that can be afforded at the PU.
Finally, we have introduced a cross-layer design of SpSe and CSMA/CA from which we have derived the transmission and collision probabilities. We rely on discrete time Markov chain model with a state pair for modeling both SpSe and CSMA/CA processes that result in a joint stationary probability which incorporates the parameters of SpSe and CSMA/CA at the physical and the MAC layer respectively.

References

1. F. Foukalas, V. Gazis, and N. Alonistioti. "Cross-Layer Design Proposals for Wireless Mobile Networks: A Survey and Taxonomy", IEEE Communications Surveys and Tutorials, vol. 10, no. 1, pp. 70-85, First Quarter 2008.
2. F.Foukalas, G.Karetsos and L.Merakos, "Cross-layer Design in Opportunistic Spectrum Access based Cognitive Radio Networks", Special Issue on Cognitive Radio Systems, accepted for publication to Int. J. of Commun. Netw. Distr. Syst., Inderscience Publishing.
3. F. Foukalas, G. Karetsos and L.Merakos, "Spectral Efficiency of Cognitive Radio Networks under Interference Constraint and QoS Guarantees", Electrical and Computer Engineering, Elsevier Publishing, DOI: 10.1016/j.compeleceng.2011.07.005.
4. F.T. Foukalas, P.T. Mathiopoulos and G.T. Karetsos, "Joint Optimal Power Allocation and Sensing Threshold selection for SU's capacity maximization in SS CRNs", IET Electr. Lett., vol. 46, no. 20, pp. 1406-1407, Sept. 2010.
5. F. T. Foukalas, G. Karetsos and L.Merakos, "Capacity Optimization Through Sensing Threshold Adaptation for CRNs", Optimiz. Lett., Springer Publishing (DOI 10.1007/s11590-011-0345-8).
6. F. T. Foukalas, G. T. Karetsos and P.Chatzimisios, "Cross-layer Design of CMSA/CA with Spectrum Sensing in CRNs", submitted to AEU Electronics and Communications, Elsevier publishing, May 2011.
7. Kang Xin, Y.-C. Liang, A. Nallanathan, H.K. Garg and R. Zhang, "Optimal Power Allocation for Fading Channels in Cognitive Radio Networks: Ergodic Capacity and Outage Capacity", IEEE Transactions on Wireless Communications, Vol. 8, No. 2, pp. 940 - 950, February 2009.
8. Haddad, M. Hayar, A. Debbah, M., "Spectral efficiency of spectrum-pooling systems", IET Communications. 2, 6, (July 2008), 733-741.
9. Haykin, S., "Cognitive Radio: Brain-Empowered Wireless Communications", IEEE Journal on Selected Areas in Communications. 23, (Feb. 2005).
10. Zhao, Q. and Sadler, B.: 'A survey of dynamic spectrum access', IEEE Sign. Proc. Mag., 2007, 24, (3), pp. 79–89.

11. Musavian, L., and Aïssa, S.: 'Capacity and power allocation for spectrum-sharing communications in fading channels', IEEE Trans. Wirel. Commun. 2009, 8, (1), pp. 148-156.
12. Kang, X., Liang, Y.-C., Garg, H. K., and Zhang, L., 'Sensing-based spectrum sharing in cognitive radio networks', IEEE. Trans. Veh. Tech., 2009, 58, (8), p.p. 4649-4654.
13. Choi, H., Jang, K., and Cheong, Y., 'Adaptive sensing threshold control based on transmission power in cognitive radio systems', Proc. 3 rd Int. Conf. on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom), 2008, p.p. 1-6.
14. Digham, F.F., AlouIni, M.-S., and Simon, M. K., 'On the energy detection of unknown signals over fading channels', IEEE Trans. Commun., 2007, 55, (1), pp. 21–24.
15. Liang, Y.C., Zeng, Y., Peh, E.C.Y., Hoang A.T.: 'Sensing-throughput tradeoff for cognitive radio networks', IEEE Trans. Wirel. Commun., 2008, 7, (4), pp. 1326-1337.
16. Cormio, C., Chowdhury, K.R.: 'A Survey on MAC protocols for cognitive radio networks', Elsevier, Ad-Hoc Netw.7, 2009, pp. 1315–1329.
17. Liu, Q., Zhou, S., Giannakis, G.B.: 'Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design', IEEE Trans. Wirel. Commun., 2005, 4, (3), pp. 1142-1153.
18. Bianchi, G.: 'Performance analysis of the IEEE 802.11 distributed coordination function', IEEE Journ. Sel. Ar. Commun., 2000, 18 (3), pp. 535–547.
19. Xing, Y., Chandramouli, R., Mangold, S., Shankar S. N.: 'Dynamic Spectrum Access in Open Spectrum Wireless Networks', IEEE Journal Sel. Ar. Comm, vol. 24, no. 3, March 2006.
20. Chong, J.W., Sung, D.K., Sung, Y.: 'Cross-Layer Performance Analysis for CSMA/CA Protocols: Impact of Imperfect Sensing', IEEE Trans. Veh. Tech., vol. 59, no. 3, March 2010.
21. Chen, Y., Zhao, Q., Swami, A.: 'Joint Design and Separation Principle for Opportunistic Spectrum Access in the Presence of Sensing Errors', IEEE Trans. Inf. Theory, 2008, 54, (5), pp. 2053-2071.

Text Comprehension Web-based Adaptive Environment for Distance Learning – Exploitation in Computer Science Education

Alexandra Gasparinatou[1]

National and Kapodistrian University of Athens

Department of Informatics and Telecommunications

alegas@di.uoa.gr

Abstract. This dissertation contributes to the field of learning in the domain of Computer Science. We investigated the effects of background-knowledge and text-cohesion on learning from texts in Computer Science. Our results showed that students with low background knowledge appeared to benefit from a high-cohesion text, whereas students with high background knowledge from a low-cohesion text. Based on our results, we designed and developed the Adaptive Learning Models from texts and Activities (ALMA) environment which supports the processes of learning and assessment via: 1) texts differing in local and global cohesion for students with low, medium and high background knowledge, 2) activities corresponding to different levels of comprehension which prompt the student to practically implement different text-reading strategies, with the recommended activity sequence adapted to the student's learning style, 3) an overall framework for informing, guiding and supporting students in performing the activities, 4) individualized support and guidance according to student-specific characteristics. ALMA also, supports students in distance learning or in blended learning in which students are submitted to face-to-face learning supported by computer technology. The adaptive techniques provided via ALMA are: a) adaptive presentation and b) adaptive navigation. Digital learning material, in accordance to the text comprehension model described by Kintsch (1998), was introduced into the ALMA environment. The material includes texts of varying local and global cohesion and activities corresponding to different comprehension levels and appropriate for all learning styles. This material can be exploited in either distance or blended learning.

1 Introduction

Learning from texts is a complex process and till now, not completely understood [1,2]. In order to optimize learning, should one make the comprehension process as easy as possible, or should one, as many educators insist, ensure that the learner participates actively and intentionally in the process of constructing the meaning of a text [3]? Specifically, should the readers' task be facilitated by improving the comprehensibility of a text or should the readers' active involvement be increased by placing obstacles in their way? In the second case, what sort of obstacles will have beneficial effects on learning and under what conditions? The approach to this question has been the study of characteristics of the text, the characteristics of the individual reader and how these factors affect text comprehension.

A considerable number of empirical studies have been conducted in order to answer this question. Many of them have demonstrated that readers' background knowledge facilitates and enhances comprehension and learning [4]. These studies have also shown that readers with greater background knowledge express more interest in the reading material and employ more effective reading strategies. Additionally, experts tend to put more effort into learning than do novices [5]. Text comprehension can also be facilitated and enhanced by rewriting poorly written texts in order to be more

[1] Dissertation advisor: Maria Grigoriadou, Professor

cohesive and to provide the reader with all the information needed for a good comprehension [6,7, 8]. Text coherence refers to the extent to which a reader is able to understand the relations between ideas in a text. This is generally dependent on whether these relations are explicit in the text.

Nevertheless, a cohesive text representation does not always result in better learning. Readers with appropriate knowledge do not always employ that knowledge for learning. They also tend to take the path of least resistance and if they have the feeling that they are easily understanding the text they read, they may not bother to activate their knowledge and form the links between it and the text that guarantee learning. Thus, there exists an instructional need to stimulate reader activity [3]. Consequently, the advantages found for facilitating the reading process by making text more cohesive and the disadvantages demonstrated for facilitating the learning process present contradictory findings.

According to Kintsch, there is no text comprehension that does not require the reader to apply knowledge: lexical, syntactic and semantic knowledge, domain knowledge, personal experience and so on. Ideally a text should contain the new information a reader needs to know plus just enough old information to allow the reader to link the new information with what is already known. Texts that contain too much that the reader already knows are boring to read and, indeed, confusing (e.g., legal and insurance documents that leave nothing to be taken for granted). Consequently, too much coherence and explication may not necessarily be a good thing.

The way in which cohesion manipulations influence the comprehension and consequently the learning from computer science texts (e.g. Computer Networks texts) may differ from that of social and natural sciences texts. Thus, it is of great importance to investigate learning from texts in Computer Science. Our study contributes to the eld of learning from texts in the domain of Computer Science. Based on previous research in various other domains, we further examined the effects of background knowledge on learning from high- and low-cohesion texts in Computer Science. Speci cally, we investigated the learning from texts in "Local Networks Topologies" by undergraduate low- and high-knowledge students.

For this purpose, we conducted three empirical studies [9,10]: (1) The main purpose of the 1st study was to investigate the effects of text-cohesion on low and high-background knowledge students learning in Computer Science. We used texts concerning the domain of "Local Networks Topologies". Participants with low- and high-background knowledge about the domain of the text were included. They were separated randomly in four groups (A, B, C, and D). Each group was given one from four text versions of different cohesion. Tasks differentially sensitive to textbase and situation model constructions were used. The participants were tested on memory recall and deep comprehension, (2)The main purpose of the present study was to investigate the effects of text-cohesion on high-knowledge students learning in computer science. We used texts concerning the domain of "Local Networks Topologies", (3)The purpose of the 3rd study was twofold. Firstly, to assess reading comprehension of students with high-and low background knowledge using texts in Computer Science with low- and high local cohesion. Next, to examine how question format (multiple-choice vs. open-ended) influence the assessment of science text comprehension among undergraduate students. We attempted to compare multiple-choice and open-ended question by directly transforming comprehension questions from one format to the other.

According to (ACM and IEEE) [11]:

- Computer science texts are complex depending on factors mainly inherent in the texts. Much of their content is abstract and technical, far removed from everyday experience.
- Computer science texts support students to utilize concepts from many different elds. All computer science students must learn to integrate theory and practice, to recognize the importance of abstraction, and to appreciate the value of good engineering design.
- Computer science texts support students to understand the theoretical underpinnings of the discipline and also how that theory influences practice.
- Computer science texts support students to develop a high-level understanding of systems as a whole. This understanding must transcend the implementation details

of the various components to encompass an appreciation for the structure of computer systems and the processes involved in their construction and analysis.
- Computer science texts must help students to encounter many recurring themes such as abstraction, complexity, and evolutionary change. They will also encounter principles, e.g. those associated with caching, (e.g. the principle of locality), with sharing a common resource, with security, with concurrency, and so on.

2 The Construction-Integration Model

The construction-integration model is an extension of earlier comprehension models [12, 13], primarily specifying computationally the role of prior knowledge during the comprehension process. It distinguishes several different levels that readers construct during the mental representation of a text. *Text base* and *situation model* understanding are most relevant for the objectives of this study. The text base contains the information that is directly expressed in the text, organized and structured in the same way as by the author. It has a local and global structure (micro- and macro-structure respectively). Micro-structure refers to local text properties, macro-structure to the global organization of text. The situation description constructed by the learner on the basis of a text, as well as prior knowledge and experience, is called the *situation model*.

Text Cohesion

The degree to which the concepts, ideas and relations with a text are explicit has been referred to as *text cohesion,* whereas the effect of text cohesion on readers' comprehension has been referred to as *text coherence* [14]. Text coherence refers to the extent to which a reader is able to understand the relations between ideas in a text and this is generally dependent on whether these relations are explicit in the text.

The Measurement of Learning

Some measures are more indicative of text memory (e.g. recognition, text-based questions, and text-recall) whereas other measures are more sensitive to learning (e.g., bridging- inference questions, recall elaborations, problem-solving tasks, keyword sorting tasks). The former are referred to as text base measures because a cohesive text base understanding is all that is required for a high performance. The latter are referred to as situation model measures because, in order to perform well, the reader must have formed a well-integrated situation model of the text during the comprehension process [3].

3 The 2nd Empirical Study-Results

3.1 Reading Rates

The time required for each participant to read the text, was recorded. The number of words in each text was divided by the reading time yielding the average number of words per minute. Participants read the text twice, yielding two reading rate scores. The results are presented in Tables 1a and 1b.

Readers read the text much more slowly the first time (M=108 words/minute) in relation with the second (M=159 words/min). A significant main effect was obtained for local cohesion, ($F(1,61) = 16.608$, $p<0.001$, for the 1st reading and $F(1,61)= 14.259$, $p<0.001$, for the 2nd reading). Students who read the texts with the maximum local cohesion (L) had higher reading rates scores (M =132 words/min for the 1st reading and M=196 words/min for the 2nd reading) than students who read the texts with the minimum local cohesion (l) (M=85 for the 1st reading and M=123 for the 2nd reading). This was a low difference (Partial Eta Squared=0.214 for the 1st reading and Partial Eta Squared=0.189 for the 2nd reading). This result indicates that the minimally cohesive text at the local level requires more inferences than does the high cohesion text.

Table 1a: The mean (standard deviation) of reading rates in words per minute

Local cohesion	Global cohesion	N	1st reading		2nd reading	
			Mean	SD	Mean	SD
Low	low	16	87	42	141	61
	high	16	84	41	104	55
	Total	*32*	*85*	*41*	*123*	*58*
High	low	16	131	39	201	97
	high	17	132	58	191	92
	Total	*33*	*132*	*49*	*196*	*94*
Total	low	32	107	46	168	83
	high	33	110	56	150	87
	Total	*65*	*108*	*51*	*159*	*85*

Table 1b: Reading rates: Tests of Between-Subjects Effects

Source	1st reading			2nd reading		
	F	Sig.	Partial Eta Squared	F	Sig	Partial Eta Squared
Local cohesion	16.608	0.000	0.214	14.259	0.000	0.189
Global cohesion	0.005	0.943	0.000	1.474	0.229	0.024
Local cohesion* global cohesion	0.036	0.851	0.001	0.491	0.486	0.008

Participants also read more quickly texts with the minimum global cohesion but there was not obtained a significant main effect $F(1,61)=0.005$, $p=0.943$, for the 1st reading and $F(1,61)= 1.474$, $p=0.229$ for the 2nd reading) indicating that students who read the texts with the high global cohesion had about the same reading rates scores (M=107 words/min for the 1st reading and M=168 words/min for the 2nd reading) with the students who read the texts with the low global cohesion (M=110 words/min, for the 1st reading and M=150 words/min for the 2nd reading) (Partial Eta Squared=.000, for the 1st reading and Partial Eta Squared=0.024 for the 2nd reading). Thus, the absence of an explicit macrostructure in the text did not slow high-knowledge participants down.The interaction effect between local and global cohesion was not statistically significant ($F(1,61)=0.036$, $p=0.851$, Partial Eta Squared=0.001, for the 1st reading and $F(1,61)=0.491$, $p=0.486$, Partial Eta Squared=0.008, for the 2nd reading) indicating that the local cohesion difference scores do not depend on the particular global cohesion (low or high). These results are consistent with McNamara et al., (1996).

3.2 Text Recall

The text paragraph concerning "Tree Topology" was propositionalized in the four text versions. In order to compare recall for the different text versions, the analysis included only those propositions containing information common to all four texts (i.e., those comprising the lg text). This scoring method allows by-item analyses to be performed because the propositions that are scored remain the same for all participants regardless of text. There were 20 micro propositions and 3 macro propositions common to all texts.

Participants recalled the text twice, once after the first and again after the second reading of the text. The two results for each participant were pooled and scored collectively. Thus, a composite recall was formed of the propositions provided in the first recall together with any additional (non repeated) propositions that occurred in the second recall. Two-way ANOVA by participants and by items was performed on proportional recall including the factors local cohesion, global cohesion and proposition type. Text recall-scores are presented in Tables 2a and 2b.

Participants reproduced texts well enough. In the recall of *micro-propositions* there was a significant main effect for local cohesion $F(1,61)=6.438$, $p=0.014$. Students who read the text with the low local cohesion had higher scores ($M=0.54$) than those who read the text with the high local cohesion ($M=0.44$). The effect size was (Partial Eta Square=.095). This result indicates that students constructed a better text base with the text of low local cohesion. In the recall of *micro-propositions*, there was not a significant main effect for global cohesion $F(1,61)=0.021$, $p=0.886$. Students who read the text with the low global cohesion had the same scores ($M=0.49$) with those who read the text with the high global cohesion ($M=0.49$). The effect size was (Partial Eta Square=.00). This result indicates that students were able to construct a good text base both with low and high global text cohesion. This result is consistent with McNamara et al., (1996). The interaction effect was not significant in the recall of micro propositions ($F(1, 61) = 0.209$, $p=0.649$). The effect size was (Partial Eta Square = 0.003) indicating that the local cohesion difference do not depend on the particular global cohesion (low or high).

Table 2a: Text-Recall scores

Local cohesion	Global cohesion	N	Proportion of Micro-propositions recalled		Proportion of Macro-propositions recalled	
			Mean	SD	Mean	SD
Low	low	16	0.53	0.14	0.67	0.44
	high	16	0.56	0.17	0.59	0.43
	Total	*32*	*0.54*	*0.15*	*0.63*	*0.43*
High	low	16	0.45	0.17	0.53	0.32
	high	17	0.44	0.16	0.62	0.45
	Total	*33*	*0.44*	*0.16*	*0.57*	*0.39*
Total	low	32	0.49	0.16	0.61	0.39
	high	33	0.49	0.17	0.60	0.43
	Total	*65*	*0.49*	*0.16*	*0.60*	*0.41*

Table 2b: Text-Recall scores: Tests of Between-Subjects Effects

Source	Recall of micro propositions			Recall of macro propositions		
	F	Sig.	Partial Eta Squared	F	Sig	Partial Eta Squared
Local cohesion	6.438	0.014	0.095	0.317	0.575	0.005
Global cohesion	0.021	0.886	0.000	0.000	0.985	0.000
Local cohesion* global cohesion	0.209	0.649	0.003	0.675	0.414	0.011

In the recall of *macro-propositions* there was not a significant main effect for local cohesion, ($F(1,61)=0.317$, p=0.575). Students who read the text with the low local cohesion had the same scores ($M=0.63$) with those who read the text with the high local cohesion ($M=0.57$). The effect size was weak (Partial Eta Square=0.005). There was not a significant main effect for global cohesion, in the recall of *macro-propositions* ($F(1,61)=0.000$, $p=0.985$). Students who read the text with the low global cohesion had about the same scores ($M=0.61$) with those who read the text with the high global cohesion ($M=0.60$). The effect size was (Partial Eta Square=0.000). The interaction effect was not significant, $F(1,61)=0.675$, $p=0.414$. The effect size was (Partial Eta Squared=0.011). These results indicate that students were able to construct a good text base both with low and high global text cohesion and are consistent with McNamara et al., (1996).

3.3 Assessment Reading Questions

Participants answered 8 open-ended questions after each of the two readings of the text. First or second assessment questionnaire completion times were combined because they were similar. There were no significant differences between the four text conditions in terms of the total amount of time spent answering questions (M=12 min, $F(3,61)=0.476$, $MSE=0.47$, $p=0.7$). The evaluation of this task was performed by the two course teachers and was expressed as percentage correct. The marking of the

performance measure was indeed blinded, i.e. the markers did not know which group each student was in. Participants answered 8 open-ended questions after each of the two readings of the text. First or second assessment questionnaire completion times were combined because they were similar. There were no significant differences between the four text conditions in terms of the total amount of time spent answering questions (M=12 min, F(3,61)=0.476, MSE=0.47, p=0.7). The evaluation of this task was performed by the two course teachers and was expressed as percentage correct. The marking of the performance measure was indeed blinded, i.e. the markers did not know which group each student was in. The results are shown in Tables 3a and 3b.

Table 3a: Proportion of correct responses to the assessment reading questions

Local cohesion	Global cohesion	N	Text-based questions		Bridging – Inference questions		Elaborative-inference questions		Problem-solving questions	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Low	Low	16	0.72	0.14	0.93	0.12	0.76	0.10	0.92	0.05
	High	16	0.65	0.16	0.78	0.21	0.66	0.19	0.75	0.17
	Total	*32*	*0.68*	*0.15*	*0.85*	*0.16*	*0.71*	*0.14*	*0.83*	*0.11*
High	Low	16	0.61	0.17	0.76	0.15	0.52	0.21	0.80	0.14
	High	17	0.70	0.17	0.88	0.15	0.68	0.18	0.74	0.13
	Total	*33*	*0.65*	*0.17*	*0.82*	*0.15*	*0.60*	*0.19*	*0.77*	*0.13*
Total	Low	33	0.67	0.16	0.85	0.16	0.65	0.20	0.87	0.12
	High	32	0.67	0.16	0.83	0.19	0.67	0.18	0.74	0.15
	Total	*65*	*0.67*	*0.16*	*0.84*	*0.17*	*0.66*	*0.19*	*0.81*	*0.13*

Table 3b: Assessment reading questions: Tests of Between-Subjects Effects

Cohesion	**Text-based questions**			**Bridging-inference questions**			**Elaborative-inference questions**			**Problem-solving questions**		
	F	Sig.	Partial Eta Squared	F	Sig.	Partial Eta Squared	F	Sig.	Partial Eta Squared	F	Sig.	Partial Eta Squared
local	0.524	0.472	0.009	0.681	0.412	0.011	6.819	0.011	0.101	4.108	0.047	0.063
global	0.049	0.826	0.001	0.114	0.737	0.002	0.424	0.518	0.007	13.583	0.000	0.182
local * global	3.300	0.074	0.051	11.646	0.001	0.160	8.653	0.005	0.124	2.252	0.139	0.036

For *text-based* question scores, there was not a significant main effect neither for local cohesion (F(1,61)=0.524, p=0.472) nor for global cohesion (F(1,61)=0.049, p=0.826). Students who read the texts with the low local cohesion had about the same scores (M=0.68) with those reading the texts with the high local cohesion (M=0.65). In addition, students who read the texts with low global cohesion had the same scores (M=0.67) with those reading the high global cohesion texts (M=0.67). The effect size for local cohesion was (Partial Eta Square=0.009) and for global cohesion was (Partial Eta Square=0.001). The interaction effect was not significant, (F(1,61)=3.300, p=0.074) indicating that the local cohesion difference scores do not depend on the particular global cohesion (low or high). The effect size was (Partial Eta Square=0.051).These results indicate that students were able to construct a good text base both with low and high local and global text cohesion and they are consistent with McNamara et al., (1996).

For *bridging-inference* question scores, there was not a significant main effect neither for local cohesion (F(1,61)=0.681, p=0.412) nor for global cohesion (F(1,61)=0.114, p=0.737). Students who read the texts with the low local cohesion had better scores (M=0.85) than those who read the texts with the high local cohesion (M=0.82) but the difference was not statistically significant. In addition, students who read the texts with low global cohesion had about the same scores (M=0.85) with those reading the high global cohesion texts (M=0.83). The effect size for local cohesion was (Partial Eta Square=0.011) and for global cohesion was (Partial Eta Square=0.002). The interaction effect was significant, (F (1, 61) = 11.646, p=0.001) indicating, although

the effect size was relatively weak (Partial Eta Square = 0.160), that the local cohesion difference scores depend on the particular global cohesion (low or high).
For *elaborative-inference question* scores, a significant main effect was obtained for local cohesion, (F(1,61)=6.819, p=0.011). Students reading texts with low local cohesion had significantly higher scores (M=0.71) than students with high local cohesion texts (M=0.60). The effect size was (Partial Eta Squared = 0.101). However, there was not obtained a significant effect for global cohesion, (F(1,61)= 0.424, p=0.518). Students reading texts with low global cohesion had about the same scores (M=0.65) with the students reading high global cohesion texts (M=0.67). The effect size was (Partial Eta Squared=0.007). The interaction effect between local and global text cohesion was significant, (F(1,61)=8.653, p=0.005), meaning that the local cohesion difference scores, depend on the particular global cohesion (low or high). The effect size was (Partial Eta Squared = 0.124).
For *problem solving* question scores, a significant main effect was obtained for local cohesion, (F(1,61)=4.108, p=0.047). Students reading texts with low local cohesion had significantly higher scores (M=0.83) than students with high local cohesion texts (M=0.77). The effect size was (Partial Eta Squared = 0.063). Additionally, a significant main effect was obtained for global cohesion, (F(1,61)= 13.583, p=0.000). Students reading texts with low global cohesion had significantly higher scores (M= 0.87) than students with high global cohesion texts (M=0.74). The effect size was (Partial Eta Squared=0 .182). The interaction effect between local and global text cohesion was not significant, (F(1,61)=2.252, p=0.139), meaning that the local cohesion difference scores do not depend on the particular global cohesion (low or high). The effect size was (Partial Eta Squared = 0.036).

3.4 Sorting Activity

Participants were randomly assigned in the four text versions. The sorting data were used to determine how strongly reading the text affected the reader's conceptual structure concerning the information in the text. We were not interested in how well or reasonably participants sort the items, but in the degree to which the information presented in the text, influences their sorting. The results of this analysis are shown in Tables 4a and 4b. The evaluation of this task was performed by the two course teachers and was expressed as percentage correct. The marking of the performance measure was indeed blinded, i.e. the markers did not know which group each student was in.

Table 4a: Proportion of correct sorted data

Local cohesion	Global cohesion	N	Pre-reading sorting activity		Post-reading sorting activity		Improvement in sorting activity scores	
			Mean	SD	Mean	SD	Mean	SD
Low	Low	16	0.78	0.15	0.94	0.11	0.16	0.11
	High	16	0.71	0.21	0.89	0.11	0.18	0.17
	Total	*32*	*0.74*	*0.18*	*0.91*	*0.11*	*0.17*	*0.14*
High	Low	16	0.69	0.21	0.75	0.22	0.06	0.02
	High	17	0.88	0.12	0.91	0.13	0.03	0.03
	Total	*33*	*0.78*	*0.16*	*0.83*	*0.18*	*0.04*	*0.02*
Total	Low	33	0.74	0.18	0.86	0.19	0.11	0.09
	High	32	0.80	0.19	0.90	0.12	0.10	0.14
	Total	*65*	*0.77*	*0.18*	*0.88*	*0.15*	*0.11*	*0.11*

Table 4b: Sorting activity: Tests of Between-Subjects Effects

Cohesion	Pre-reading sorting activity			Post-reading sorting activity			Improvement in sorting activity		
	F	Sig.	Partial Eta Squared	F	Sig.	Partial Eta Squared	F	Sig.	Partial Eta Squared
local	0.963	0.330	0.016	4.876	0.031	0.074	22.755	0.000	0.272
global	1.869	0.177	0.030	1.866	0.177	0.030	0.033	0.857	0.001
local * global	8.405	0.005	0.121	8.652	0.005	0.124	1.057	0.308	0.017

According to Tables 4a and 4b, in *pre-reading sorting activity*, a significant main effect was obtained for local cohesion, $F(1,61)=0.963$, $p=0.330$. In *post-reading sorting activity* a significant main effect was obtained for local cohesion, $F(1,61)=4.876$, $p=0.031$. Students reading the texts with low local text cohesion, had significantly higher scores ($M=0.91$) than students who read the texts of high local text cohesion ($M=0.83$). The effect size was (Partial Eta Squared = 0.074). There was not obtained a significant main effect for global cohesion, $F(1,61)=1.866$, $p=0.177$. Students reading the texts with low global text cohesion, had about the same scores ($M=0.86$) with those reading high global text cohesion ($M=0.90$). The effect size was (Partial Eta Squared=0.030). The interaction effect was also significant, ($F(1,61)=8.652$, $p=0.005$, Partial Eta Squared=0.124) indicating that the local cohesion difference scores depends on the particular global cohesion (low or high).
As we are interested in the degree to which the information presented in the text, influences students' sorting, the most important is the *improvement in sorting activity*. A significant main effect was obtained for local cohesion, $F(1,61)=22.755$, $p=0.000$. Students reading the texts with low local text cohesion, had significantly higher improvement ($M=0.17$) than students reading the texts of high text cohesion ($M=0.04$). The effect size was (Partial Eta Squared=0.272). There was not obtained a significant main effect for global cohesion ($F(1,61)=0.033$, $p=0.857$, Partial Eta Squared=0.001). The interaction effect was not significant, ($F(1,61)=1.057$, $p=0.308$, Partial Eta Squared=0.017) indicating that the local cohesion difference scores, do not depend on the particular global cohesion category (low or high). Consequently, high-knowledge readers developed a better situation model with the texts of minimum local cohesion (lg, lG). This result is consistent with McNamara et al., (1996) for high-knowledge readers and Kintsch & McNamara, (1996, 1st experiment).

3.5 Conclusions

Our study demonstrated that learners with adequate background knowledge, reading a text with minimum cohesion were forced to infer unstated relations in the text and were engaged in compensatory processing at the level of the situation model. This enabled them to understand the text more deeply than if they were given a more cohesive text. These results confirm the findings of previous studies, such as in the domain of heart disease [4]. Understanding the ways and directions in which text structure, individual differences and comprehension measures interact, is vital for a complete theoretical account of text comprehension, as well as an educational approach to using texts in a classroom. W. Kintsch's model of text comprehension has provided us with a framework to approach these issues.
This research suggests an approach in which the cohesion level of the text is adjusted to the student's level of knowledge, so that reading becomes challenging enough to stimulate active processing but not so difficult as to break down comprehension. This would mean constructing several versions of a text in order to accommodate varying levels of knowledge among readers. According to McNamara et al. [4], the idea of "customizing" a textbook is not as impractical as it may seem. For example, textbook publishers provide instructional texts that are a composite of particular subject areas requested by individual teachers. Moreover, the kind of educational application of customized text is easily within the capability of present day hypertext computer systems. Text could be presented on a computer screen with interspersed questions or tasks designed to assess a student's comprehension online. Instructional text could then be presented at the level of cohesion that is appropriate to the student's current level of understanding so that it encourages inferencing but also ensures that the reader is able to do so. In this way, students are forced to use their knowledge as they read, allowing effective learning from a textbook to be achieved by a much wider range of students than is possible with a single text targeted at a supposed average reader.
Moreover, as it concerns the assessment using open-ended questions versus the assessment using multiple choice questions, our 3rd study confirmed that comprehension assessments using open-ended questions are comparable to those using multiple-choice questions only in the case of elaborative-inference questions for both students with high- and low background knowledge. The results show that elaborative-inference questions were more difficult than text-based and bridging-

inference questions regardless of the question format. This occurs because in text-based and bridging-inference multiple choice questions with very selective and controlled distracter options might never be exactly the same as open-ended questions because multiple–choice questions provide richer retrieval cues than corresponding open-ended questions [16]. On the other hand, in order an elaborative-inference question to be answered, linking text information and information from outside knowledge is required. Thus, answering this type of question requires the integration of text information with background knowledge. In this case the distracter options do not include information which is contained entirely in the text.
Consequently elaborative-inference questions assess the situation model a student constructs during reading both in open-ended and in multiple choice format. On the other hand bridging-inference questions assess the situation model a student construct during reading only in open-ended format whereas bridging-inference questions in multiple choice format assess the text-base model as do text-based questions in both formats.

4. Text Comprehension Web-based Learning Environments

In the field of text comprehension, many researchers have been examining issues focusing on assisting comprehension through personalized learning environments. Point & Query (P&Q) is an environment where students learned entirely by asking questions and interpreting answers to questions [17]. AutoTutor, holds a conversation in natural language that coaches the student in constructing a good explanation in an answer, that corrects misconceptions, and that answers student questions [17]. MetaTutor, a hypermedia environment, designed to train and foster students' self-regulated learning (SRL) [18, 19]. ReTuDiS is a tutorial dialogue system for learner modeling text comprehension through personalized reflective dialogue [20]. Interactive Strategy Training for Active Reading and Thinking (i-START) is a web-based application that provides young adolescent to college students with high-level reading training to improve comprehension of science texts. i-START is modeled after an effective, human-delivered intervention called self-explanation reading training (SERT), which trains readers to use active reading strategies to self-explain difficult texts more effectively [21, 22].

5. An Outline of the ALMA Environment (Adaptive Learning Models from Texts and Activities)

Fig. 1: An Outline of the ALMA Environment

ALMA [23] actively engages students in the learning process. It takes into account readers' background knowledge in order to propose the appropriate text version from four versions of a text with the same content but different cohesion at the local and

global level. To achieve this goal, it suggests that the student performs a background knowledge assessment test, with scores characterized as "high", "median" and "low". ALMA motivates high knowledge students to read the minimally cohesive text at both local and global levels (lg), median knowledge students to read the text with maximum local and minimum global cohesion (Lg) or with minimum local and maximum global cohesion (lG) and low knowledge students to read the maximally cohesive text (LG). ALMA also allows the student to choose the preferred version of text and records the time spent reading it. The following three types of rules were used to maximize local cohesion: (1) replacing pronouns with noun phrases when the referent was potentially ambiguous (e.g. In the phrase: *"This has been very popular for exchanging music files via the internet"*, we replace *"This"* by *"The peer-to-peer model"*. (2) Adding descriptive elaborations linking unfamiliar and familiar concepts (e.g., *"In networks, computers users can exchange messages and share resources",*is elaborated to: *"In networks, computers users can exchange messages and share resources-such as printing capabilities, software packages, and data storage facilities-that are scattered throughout the system").* (3) Adding sentence connectives (however, therefore, because, so that) to specify the relation between sentences or ideas. In the global macro cohesion versions of the texts (lG and LG), macro propositions were signaled explicitly by various linguistic means (i.e., macro signals): (1) adding topic headers (e.g., Network Classifications, Protocols) and (2) adding macro propositions serving to link each paragraph to the rest of the text and the overall topic (e.g., *"Afterwards, the rules by which network activities are conducted, will be discussed"*) [4].

ALMA supports and assesses students' comprehension through a series of activities such as: text recall, summaries, text-based, bridging inference, elaborative inference, problem solving, case studies, active experimentation and sorting tasks. *Text recall,* helps students remember the basic ideas in the text by translating it into more familiar words. The students are also encouraged to go beyond the basic sentence-focused processing by linking the content of the sentences to other information, either from the text or from the students' background knowledge. The empirical findings have shown that students who are able to recall the text and go beyond the basic sentence-focused processing are more successful at solving problems, more likely to generate inferences, construct more coherent mental models, and develop a deeper understanding of the concepts covered in the text [20]. *Summaries* also encourage students to go beyond the text and like text recall can be perfectly good indicators of well-developed situation models [3]. *Text-based questions,* as they demand only a specific detail from the text, measure text memory. *Bridging-inferences questions* motivate students to make *bridging inferences* which improve comprehension by linking the current sentence to the material previously covered in the text [21]. Such inferences allow the reader to form a more cohesive global representation of the text content (Kintsch, 1998). *Elaborative-inference questions* motivate students to associate the current sentence with their own related background knowledge. The most important is that students are encouraged to engage in logical or analogical reasoning process to relate the content of the sentence with domain-general knowledge or any experiences related to the subject matter, particularly when they do not have sufficient knowledge about the topic of the text. Research has established that both domain knowledge and elaborations based on more general knowledge are associated with improving learning and comprehension [22]. *Elaborations* essentially ensure that the information in the text is linked to information that the reader already knows. These connections to background knowledge result in a more coherent and stable representation of the text content [3, 4]. *Problem-solving questions* motivate students to use the information acquired from the text productively in novel environments. This requires that the text information be integrated with the students' background knowledge and become a part of it, so that it can support comprehension and problem solving in new situations [3]. *Sorting task* has great potential as a simple task and can be used both as a method of assessment and as a mode of instruction. Students are asked to sort a set of key words contained and not contained in the text, in certain groups. They are encouraged to do this task twice, once before reading the text and once more after reading the text. The sorting data are used to determine how strongly reading the text affected students' conceptual structure concerning the

information in the text. We are interested in the degree to which the information presented in the text influences their sorting. Sorting task is an alternative method for assessing situation model understanding. *Active experimentation activities* motivate students to undertake an active role and through experimentation to construct their own internal representations for the concept they are studying. *Case studies* motivate students to engage in the solution of an authentic and thus interesting problem. They are asked to analyze it and propose solutions. The problem is described in detail and is followed by a series of questions aiming to guide the students in the problem solving procedure.

Moreover ALMA supports multiple Informative, Tutoring and Reflective Feedback Components, aiming to stimulate learners to reflect on their beliefs, to guide and tutor them towards the achievement of specific learning outcomes and to inform them about their performance [25]. ALMA also actively engages students in the learning process by taking into account readers' learning preferences in order to propose them to start from activities that match their learning preferences and continue with less "learning preferences matched" activities in order to develop new capabilities [26]. To achieve this goal, it suggests that the student performs the "Learning-Style Inventory (LSI *©1993 David A. Kolb, Experience-Based Learning Systems,Inc.)".* The Learning – Style Inventory describes the way a student learns and how he/she deals with ideas and day-to-day situations in his/her life. It includes 12 sentences with a choice of endings. Consequently, ALMA is adapted to students' background knowledge and learning style resulting in personalized learning.

ALMA also includes the authoring tool (ALMA_auth). This tool provides the author with the option of developing and uploading the educational material. Finally, ALMA includes a *forum* where students have the possibility to collaborate with each other and also with the instructor.

6. The Assessment of ALMA environment

The Empirical Study demonstrated that ALMA could be a valuable tool for supporting the learning process in introductory computer science courses and helping students to deepen their understanding in the undergraduate curricula of computer science. Students had a positive opinion about ALMA environment because they were activated to use their background knowledge while reading and they believe that ALMA gives the opportunity to achieve better results in learning from texts in computer science than reading a single text target at an average reader. Moreover, students had a positive opinion about the learning sequence proposed by ALMA and they believe that a combination of the traditional teaching method and ALMA environment would be the best for their under and postgraduate studies. The assessment of ALMA demonstrated that the ALMA environment satisfactorily supported the learning process of students in Computer Science and almost all its functions are useful and user-friendly.

References

1. Fletcher, M. J. (2006). Measuring reading comprehension. *Scientific Study of Reading,* 10, 323-330.
2. Snow, C.E. (2003). Assessment of reading comprehension. In A.P. Sweet & C.E. Snow (Eds.), *Rethinking reading comprehension* (pp.192-218). New York: Guilford.
3. Kintsch, W. (1998). *Comprehension. A paradigm for cognition.* Cambridge: Cambridge University Press.
4. McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14,* 1-43.
5. Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64,* 37-54.
6. Beyer, R. (1991). Psychologische Untersuchungen zur Gestaltung von Instruktionstexten [Psychological studies concerning the construction of instructional texts]. *Mathematisch-Naturwissenschaftliche Reihe, 39,* 69-75. (Scientific journal published by Humboldt University, Berlin).
7. Britton, B.K., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inferences calls on recall and cognitive structures. *Journal of Educational Psychology, 83,* 329-345.

8. Mckeown, M.G., Beck, I.L., Sinatra, G.M., & Loxterman, J.A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly, 27,* 79-93.

9. Gasparinatou, A., & Grigoriadou, M. (2010). Learning from Texts in Computer Science. *The International Journal of Learning, 17*, Issue 1, ISSN: 1447-9494, pp.171-189.

10. Gasparinatou, A., & Grigoriadou, M. (2011). Supporting students' learning in the domain of Computer Science. *Computer Science Education*, *21*, ISSN: 0899-3408, pp. 1-28.

11. ACM and IEEE, (2008). Computer Science. Curriculum 2008: An Interim Revision of CS 2001. Report from the Interim Review Task Force. December 2008, published by the Association for Computing Machinery and the IEEE Computer Society.

12. Kintsch, W., & van Dijk, T.A. (1978). Towards a model of text comprehension and production. *Psychological Review, 85,* 363-394.

13. Van Dijk, T.A., & Kintsch, W. (1983). Strategies of discourse comprehension. San Diego, CA: Academic Press.

14. Graesser, A.C., McNamara, D.S., & Louwerse, M.M. (2003). What Do Readers Need to Learn in Order to Process Coherence Relations in Narrative and Expository Text? In A. P. Sweet & C.E. Snow (Eds.), *Rethinking Reading Comprehension* (pp. 82-98). New York: Guilford Publications.

15. McNamara, D.S., & Kintsch, W. (1996). Learning from texts: effects of prior knowledge and text coherence. *Discourse Processes, 22,* 247-288.

16. Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D.S. (2007). Influence of question format and text availability on assessment of expository text comprehension. *Cognition & Instruction, 25,* 399-438

17. Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist, 40*, 225–234.

18. Azevedo, R. (2009). The role of self-regulation in learning about science with hypermedia. In D. Robinson & G. Schraw (Eds.), *Current perspectives on cognition, learning, and instruction*.

19. Azevedo, R. (2008). The role of self-regulation in learning about science with hypermedia. In D. Robinson & G. Schraw (Eds.), *Recent innovations in educational technology that facilitate student learning* (pp. 127-156). Charlotte, NC: Information Age Publishing.

20. Grigoriadou M., Tsaganou G., Cavoura Th.(2005). Historical Text Comprehension Reflective Tutorial Dialogue System. *Educational Technology & Society Journal*, Special issue, 8(4), pp. 31-41.

21. Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.

22. Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology, 54,* 31–39.

23. Gasparinatou, A., & Grigoriadou, M. (2011). ALMA: An Adaptive Learning Models environment from texts and Activities that improves students' science comprehension. *Procedia-Social and Behavioral Sciences Journal, Vol. 15*, p.2742-2747 , ELSEVIER.

24. Pressley, M.,Wood, E.,Woloshyn, V., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychology, 27,* 91–109.

25. Gouli, E., Gogoulou, A., Papanikolaou, K., & Grigoriadou, M. (2006). An Adaptive Feedback Framework to Support Reflection, Guiding and Tutoring. In G.Magoulas and S.Chen (Eds.) Advances in Web-based Education: Personalized Learning Environments, 178-202.Idea Group Inc. ISBN:1-59140-691-9.

26. Kolb DA (2000).*Facilitator's guide to learning*. Boston: Hay/McBer.

Cooperative mechanisms for information dissemination and retrieval in networks with autonomous nodes

Eva Jaho*

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
Ilissia, 157 84 Athens, Greece
ejaho@di.uoa.gr

Abstract. In this dissertation we propose, model and evaluate novel algorithms and schemes that allow information dissemination and retrieval to be performed more efficiently in a modern networking environment. Most of these schemes are examined in relation to the content management tasks of content storage and classification. An inherent challenge lies in the need to manage the autonomy of nodes while preserving the distributed, as well as open nature of the system. To this end, we examine in some cases the development of incentives for nodes to cooperate while performing communication tasks. Finally, a novel attribute of most of the proposed schemes is the exploitation of social characteristics of nodes, focusing on how common interests of nodes can be used to improve communication efficiency.

1 Introduction

The proliferation of networking technologies in our time, due mostly to the development of the Internet, cellular and ad-hoc wireless mobile networks, is associated with a huge increase in the type and volume of information transmitted through a telecommunications network. Apart from the development of new communication techniques, the number of information sources has increased significantly, with user devices able to create, reproduce and transmit content that can be interesting and useful to other users. Algorithms for information dissemination and retrieval must adapt to this new setting by exploiting the ability of user devices to communicate directly and to obtain content from different sources, while handling the problem of information explosion caused by the abundance of data.

This thesis contributes to the literature by proposing, modeling and evaluating novel information dissemination and retrieval algorithms and schemes that can render communication more efficient in a modern networking environment. Such an environment consists predominantly of mobile nodes that obtain or exchange content with other peer nodes, fixed Internet servers or sensor nodes.

* Dissertation Advisor: Ioannis Stavrakakis, Professor

Apart from information dissemination and retrieval, related content management functions we examine are content storage, replication and classification.

1.1 Outline of the thesis

We provide an outline of the thesis, referring to the publications that were produced during our study. In [1] and [2] we study a content replication scheme in which autonomous nodes form a group, called a distributed replication group and cooperate in order to effectively retrieve information objects from a distant server. Each node locally replicates a subset of the server objects and can access objects stored by other nodes in the group at a smaller cost, compared to the cost of accessing them from the server. Given that nodes are autonomous and independently decide which objects to replicate, the problem is to construct efficient distributed algorithms for content replication that induce low overall average access cost. This problem becomes even more challenging when the group has to deal with churn, i.e., random "join" and "leave" events of nodes in the group; churn induces instability and has a major impact on cooperation efficiency. Given a probability estimate of each node being available, we propose a distributed churn-aware object placement strategy. By considering a game-theoretic approach, we identify cases where the churn-aware strategy is individually rational for all nodes, while the churn-unaware is not. Numerical results further show that the algorithm outperforms, in most cases, its churn-unaware counterpart, and allows for a more fair treatment of nodes according to their availability frequency, thus inciting nodes to cooperate.

Based on this setting, in [3] and [4] we study the impact of the similarity in nodes' preferences or interest profiles on content replication. Our aim is to investigate the kind of content placement strategy a node participating in a distributed replication group should follow in order to increase its benefits. We define a metric that captures the similarity of nodes' interest profiles, called group tightness. Using this metric and testing with different interest profiles, we are able to show the association of the degree of interest similarity within nodes in a group with the benefits they incur by applying a cooperative or selfish replication strategy.

An important as well as anticipated conclusion is that the higher the interest similarity between nodes, the higher the gains by cooperation in content management. It is therefore reasonable to attempt to organize nodes into communities where nodes share similar interests. However, in current "computerized" social networks users do not necessarily create ties based on common interests, but also on many other factors, such as friendship, kinship, professional relations, or even prestige. The result is a relatively small tightness of such groups, and relatively poor gains from cooperation. In [5], we propose a framework for the construction of communities based on common interests of users, by building a virtual graph where an edge between two nodes is weighted by the degree of similarity in their interests, and then using known community detection algorithms to establish communities. Testing on synthetic network scenarios shows that this framework helps to correctly identify interest communities, stressing that care

must be taken on the proper choice of the similarity metric used to represent weights.

Besides common interests, a major characteristic of communities is their locality, i.e., the specific neighbourhood, venue, or spot where they are located. Some localities may constitute points of attraction or hotspots with a higher node density. Mobile nodes also form social groups dynamically, as they move to different localities where they can establish communication with other nodes. Social groups that are examined from the viewpoint of the locality they are situated in, are termed locality-induced groups. In [6], we investigate the intermingling of interest and locality-induced social groups and propose an approach that can enhance content dissemination in the presence of such groups. We assume a setting where nodes have different interest distribution patterns over a set of information objects, and different frequencies of visiting a number of localities. Considering a new metric for the valuability of content, that takes into account both its usability and discoverability, we explore the conditions under which a cooperative strategy can improve the content dissemination process compared to a selfish one.

We further investigate content dissemination under node mobility in [7], by considering a so-called nomadic sensor network consisting of: a) sensor nodes (T-nodes), that are fixed at some points and collect information about states or variables of the environment, and b) mobile nodes (U-nodes) that collect and disseminate this information. Mobile nodes are assumed to be interested in different subsets of sensor node information. Similarly to custom multi-hop forwarding, dissemination of information content in such networks can be achieved at smaller costs if mobile nodes are cooperative and collect and carry information not only in their own interest, but also in the interest of other mobile nodes. A specific modeling scenario is considered where mobile nodes move randomly on a graph, collecting information from fixed nodes located at the vertices. We present a game-theoretic analysis to find conditions under which a cooperative equilibrium can be sustained.

Finally, in [8], we study gossip-based algorithms for content dissemination and search in large-scale networks with autonomous nodes. The dissemination or search process is carried out in rounds, where at each round multiple peers can be contacted. We develop an analytical model that allows us to evaluate the performance of the algorithm as well as the impact of several design parameters, such as the degree of cooperation of nodes, the number of peers contacted in each round, or the number of nodes where a searched content may be located. We also consider the degree of information a node has about the evolution of the gossiping process, meaning the number of nodes contacted so far, and study both the case where a node has complete information and the case of no information. The results provide significant insights on the design of such schemes.

1.2 Related works

Properly distributing replicas of content in multiple nodes in the network has shown to provide smaller search and retrieval times for content in peer-to-peer

networks, and to decrease the overall network load [9]. In mobile networks, there has been in recent years a tremendous increase in the volume of downloaded data from the Internet, which may result in congestion in wireless access links. Replication techniques have also been proposed in this case to take advantage of device-to-device communication capabilities, reduce content retrieval times and mitigate congestion in Internet access links [10], [11], [12]. Despite the fact that in a mobile network the topology dynamically changes over time, there has been significant evidence that non-random clustered mobility characterizes human movements in outdoor environments [13]. That is, despite node mobility, there is a tendency for the formation of groups composed of nodes which are in geographical proximity for a relatively long period of time and have high connectivity. This is the key fact that allows replication strategies to be extended to such networks, since it allows nodes to rely on other nodes in order to retrieve content.

A basic model we employ in our work is the one introduced in [14], where nodes are self-organized into what we call a "replication group", i.e., a group consisting of nodes in network proximity, where the cost of each node to retrieve locally stored content from another node in the group is about the same and small compared to the cost of retrieving it from the origin server. Further, we use as a basis the work in [15], where the authors devised a cooperative content placement strategy on the basis of game-theoretic arguments, determining which objects each node should store locally so that the gain for each and every node is at least (and typically much higher than) that induced under a selfish strategy.

For some content dissemination scenarios, it was demonstrated in [16] that higher similarity in the interests/preferences of online social group members favors collaborative, and even altruistic, behavior. In order to examine if such similarity is present in social networks, we devise mechanisms and tools that can assess the similarity of interests among social group members and discover interest-based communities. In the literature, algorithms for detecting community structure have largely been applied to a given network structure, usually modeled as a graph. The most prominent algorithm thereof is that of Girvan and Newman [17], which is highly efficient and overcomes many shortcomings of previously proposed algorithms. In the thesis, we shall see that interest-based relationships in social networks can be represented in the form of weighted graphs. We use a similar algorithm for detecting communities in weighted graphs [18], based on a simple mapping from a weighted network to an unweighted multigraph.

There has also been some research on how locality-induced node encounters and the nodes' own (content) interests can be jointly exploited to improve information dissemination in social networks. In this respect, the closest work to ours is in [19], where the authors introduce a dynamic scheme for deciding which objects (content) of a certain content type to replicate locally based on the encounters with other nodes. In that work each node appends a value to each object that is a function of its access probability and its availability in a locality,

its size and the weight of the locality; this weight represents the relationship between the node and the locality (e.g., how often a node visits this locality).

A nomadic sensor network is a networking paradigm that was introduced in [20]. Compared to traditional sensor networks where communication to end-users is realized in a multi-hop fashion, this paradigm exploits user mobility to conserve limited sensor energy, prolonging the lifetime of the network and making it more cost-efficient. Previous applications of game-theoretic methods in examining cooperation between mobile nodes have focused mainly on traditional ad hoc networks, where cooperation consists of each node acting as a relay and forwarding packets of other nodes, at the expense of an increased processing and energy cost. This kind of cooperation is the main subject of the papers in [21], [22], [23], [24]. A work with a similar subject to ours is [25]. Therein, the authors consider a general delay-tolerant network where information is disseminated in a store-carry-and-forward manner.

Finally, the term "gossiping algorithm" encompasses any communication algorithm where messages between two nodes are exchanged opportunistically, with the intervention of other nodes that act as betweeners or forwarders of the message. Attractive characteristics of gossiping algorithms include simplicity, scalability and robustness to failures, as well as a speed of dissemination that is easily configurable. Gossiping can be identified with the spreading of rumors in a network, the dynamics of which are investigated in [26], [27]. The process of communication consists of one or more rounds, in which a number of nodes that carry the message contact their peers, until the message reaches the intended recipient(s). In this respect, Pittel [26], Karp et al. [28] and Kempre et al. [29] studied a simpler model in which each peer selects a single neighbour in the network to communicate with at every round.

2 Results and Discussion

In the content replication scenario, let $\mathcal{N} = \{1, 2, ..., N\}$ denote the set of the nodes (or players) in a replication group and let $\mathcal{M} = \{1, 2, ..., M\}$ denote the set of objects (or items) these nodes are interested in. Let R_m^n denote the preference rate of node n for object m, and let $R^n = \{R_1^n, R_2^n, ..., R_M^n\}$ be the global preference vector. By P_n, we denote the placement at node n, defined to be the set of objects stored locally at that node with storage capacity C_n. Let t_l, t_r and t_s denote the cost for accessing an object from the node's local memory, from another remote node within the replication group and from nodes outside the replication group or distant server, respectively; $t_l < t_r < t_s$. Given an object placement $\mathcal{P}$, and node churn data, expressed in the form of ON probabilities $\pi_k, k = 1, ..., N$, $k \neq n$, the mean access cost incurred to node n per unit time

for accessing its requested objects is given by:

$$\mathcal{C}_n(P) = \sum_{i \in P_n} R_i^n t_l + \sum_{\substack{i \notin P_n, \\ i \notin P_{-n}}} R_i^n t_s + \sum_{\substack{i \notin P_n, \\ i \in P_{-n}}} \left[R_i^n [t_r (1 - \prod_{\substack{k=1, \\ k \neq n, k: i \in P_k}}^{N} (1 - \pi_k)) + t_s \prod_{\substack{k=1, \\ k \neq n, k: i \in P_k}}^{N} (1 - \pi_k)] \right]. \quad (1)$$

In the churn-aware strategy proposed in [1] and [2], each node n stores objects with the aim to minimize its access cost shown by (1), given the placements of other nodes. We show that in the majority of test cases, this strategy decreases both individual and total access costs of nodes in a replication group, compared to the strategy which does not consider other nodes' placements (called "greedy local") or the strategy which considers other nodes' placements but is unaware of churn (called "churn-unaware strategy").

An example of the importance of churn-awareness in such a distributed algorithm is shown in Fig. 1. Consider two nodes, node 1 and node 2, with capacities $C_1 = 4$, $C_2 = 1$ and 5 distinct objects $\{1, 2, 3, 4, 5\}$. Nodes 1 and 2 have corresponding request rates[1] $R^1 = \{0.5, 0.4, 0.3, 0.2, 0.1\}$ and $R^2 = \{0.4, 0.3, 0.5, 0.2, 0.1\}$. Node 1 is a relatively reliable node with $\pi_1 = 0.9$ and node 2 has a variable probability between 0 and 1 for the purpose of the example. We assume that $t_l = 1$, $t_r = 10$ and $t_s = 100$.

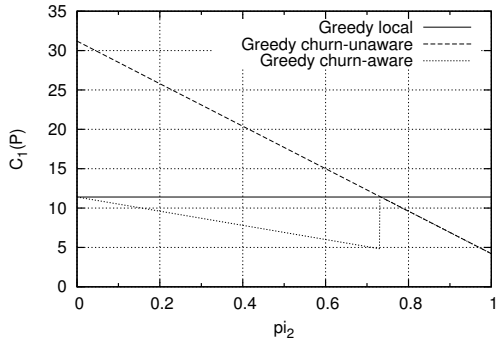


Fig. 1. Violation of the participation constraints of node 1 for $\pi_2 < 0.74$ when using the churn-unaware strategy.

Suppose that node 1 plays first in the game. It is shown in the results of Fig. 1, that for $\pi_2 < 0.74$ the access cost of node 1 under the churn-unaware strategy is greater than that of the greedy local strategy, and thus the churn-unaware strategy is not individually rational (i.e., node 1 is mistreated). This happens

[1] It is noted that request rates in the examples used here are not normalized.

because node 1 erroneously considers node 2 to be reliable. On the contrary, the churn-aware strategy always performs better than the greedy local one.

By easier satisfying individual rationality constraints and providing fairer access costs to nodes according to their reliability, the churn-aware strategy incites nodes to collectively store data. The churn-aware and churn-unaware strategies are also called "self-aware cooperative" in the thesis, as they are a mixture of selfish and cooperative behaviour.

In [3] and [4], we examined the impact of the level of similarity between node preferences within a social group on the replication strategy that nodes follow. The definition of *tightness* draws on the symmetrized Kullback-Leibler (KL) divergence [30], a well-known measure of divergence between two distributions. The Kullback-Leibler divergence of distribution R^i from R^j is defined as:

$$D_{R^i,R^j} = \sum_m R^i_m log \frac{R^i_m}{R^j_m}.$$

and its symmetrized counterpart is $D(R^i||R^j) = D_{R^i,R^j} + D_{R^j,R^i}$. Finally, we define *tightness* T to be the inverse of the average divergence of nodes' preferences within the group:

$$T = \frac{1}{\frac{\sum_{(i,j)} D(R^i||R^j)}{N(N-1)/2}}. \quad (2)$$

Our results showed that similarity is key for deciding which content placement strategy (i.e., what kind of behaviour) to follow in a replication group. We examined 3 different strategies in an environment without churn: a) the selfish (greedy local) strategy where a node stores objects based solely on its own interests, b) the self-aware cooperative strategy discussed previously, and c) the altruistic strategy which allocated object replicas in order to minimize the total access cost of the group.

In summary, our evaluation shows that the benefits of cooperation increase with the group tightness. Fig. 2 plots the total access cost values for different tightness values of the group, for a rank-preserving preference similarity scenario. Altruism emerges as a suitable strategy in very tight social groups: it minimizes the content access cost not only collectively for the whole group but also for each individual node. As tightness decreases, the collective group gain under the altruistic strategy fades out, while certain nodes may even be mistreated. For low-tightness groups, the selfish strategy is more reasonable. The performance of the self-aware cooperative strategy lies between the two; it is a suitable alternative to the altruistic strategy because it does not lead to mistreatment, and has a smaller complexity.

Based on the above results we proposed a framework called ISCoDe for the clustering of users (nodes) according to common interests. Communities in online social networks do not usually exhibit a high degree of interest similarity; thus the framework can be used as a guide for the formation of more interest-coherent communities in online social networks.

The framework is shown schematically in Fig. 3. Input to ISCoDe are the

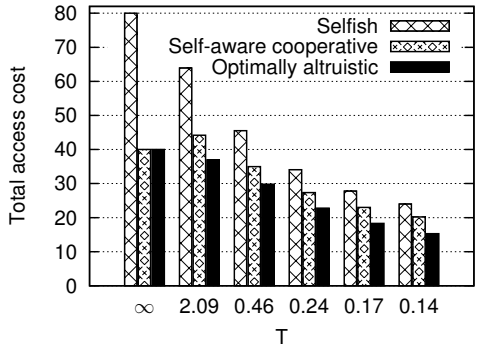


Fig. 2. Total access cost vs. *tightness* T.

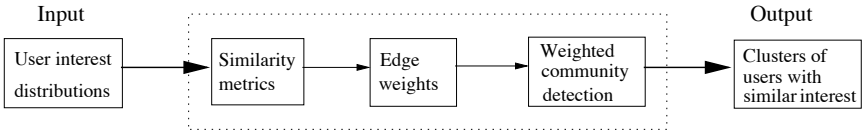


Fig. 3. The *ISCoDe* framework.

interests of the communities' member nodes in certain thematic areas. ISCoDe then proceeds in two steps. First, it quantifies the interest similarity between node pairs through the use of interest similarity metrics. Outcome of this step is a weighted graph representation of the social network, with edge weights corresponding to the similarity metric values. In a second step, ISCoDe invokes a greedy agglomerative algorithm to iteratively find the communities with the highest modularity. We have investigated two similarity metrics, Proportional Similarity (PS) and Inverse Kullback-Leibler distance (InvKL), for weighting edges according to the similarity of preferences of node-pairs in a virtual graph.

Our analytical results in the thesis suggest that both metrics produce reasonable partitions for strong community structure. However, the PS metric is more sensitive in identifying partitions in networks with less apparent interest community structure. On the other hand, InvKL has a higher resolution in networks of nodes with highly similar interests, i.e., it is able to identify smaller-sized communities.

The basic network graph model studied in [7] is as follows: U-nodes move on the graph according to a random waypoint model, with constant velocity v. Each node incurs a cost for collecting T-node content in which it is not interested. However, if all nodes are cooperative and carry content for other nodes, the benefits for all nodes outweigh the costs. We show that the following strategy of each U-node can result in a cooperative equilibrium: initially, a U-node is cooperative and copies unwanted content. However, if it meets a selfish U-node somewhere on a leg, it will only transmit its acquired content with a certain probability.

The proposed strategy may easily be applied, provided that upon meeting each other, U-nodes exchange messages that contain the list of information objects stored in their memory.

For a U-node moving on leg from T_i to T_j, the cooperative equilibrium condition is

$$\begin{aligned} N \geq & 1 + \frac{2}{\alpha} + \frac{2(1-\alpha)}{\alpha}\Big\{\frac{1}{\alpha N} - (1-p)^{N-1} \\ & - \frac{1}{aNp}\big[1-(1-p)^N\big] + (1-\alpha p)^{N-1} + \frac{cv}{d}\Big\} , \end{aligned} \tag{3}$$

where N is the number of nodes, p is the content delivery probability (same for all nodes), and α is the probability a U-node meets with another U-node before hitting T_j.

For reasonable network topology parameter, it is shown that a small number of U-nodes following this strategy (less than 10 nodes) is sufficient to sustain an equilibrium, even with a high content delivery probability.

In the study of gossiping algorithms presented in [8], we consider an initiator node I, and N other nodes on a graph. In the search case, there is a file f located in m of the other nodes of the graph ($m < N$) that the initiator wants to find. In the first round ($r = 1$) the initiator selects randomly k neighbours or gossiping targets, $1 \leq k \leq N$, to forward the message to. In each round, all the informed nodes select k gossiping targets randomly and independently to forward the message to. In the case of dissemination the objective is to inform all nodes in the network or a significant portion of them in the shortest possible time. In the case of search the query can be stopped when the object is found for the first time. A queried or informed node may or may not accept to forward the message. If it accepts, we say that this node is cooperative, otherwise non-cooperative. Cooperative nodes which are queried or informed become "active" and participate in the search or dissemination.

Some results from this analysis are shown in Fig. 4, where we plot the mean number of rounds and the mean number of active nodes until a file is found, as a function of the number of nodes in the network, N, and the cooperation probability c. We assume the search is blind, in that a node may query the same node more than once. We observe that the scaling performance of the search is remarkably simple. The mean number of rounds increases linearly with $\log N$, while the mean number of active nodes increases linearly with N. This is true for almost the whole range of values of c. Similar scaling results are derived for other cases, for example smart search (where nodes remember contacted peers) or different cooperation behaviors (e.g., stifling). The analysis also reveals interesting results for dissemination gossiping algorithms, and additionally a 0-1 bimodal behavior: the probability of informing all nodes in the network increases very abruptly from 0 to 1 after a critical round value, dependent on the size of the network.

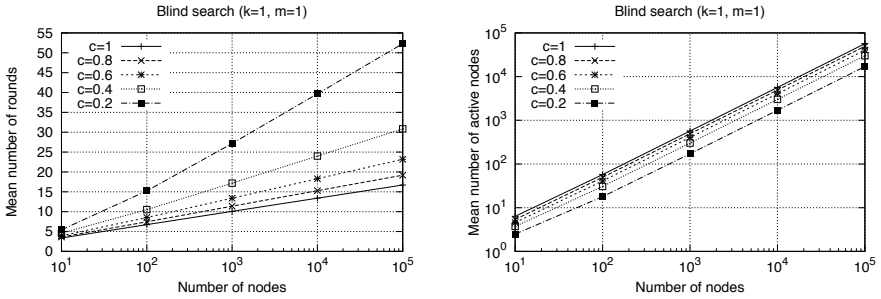


Fig. 4. Scaling performance of blind search in the plain non-cooperative case, for $k = 1$, $m = 1$: (a) mean number of rounds, (b) mean number of active nodes.

3 Conclusions

We end this summary by providing the major conclusions of the thesis.

In a distributed replication scenario, we showed that an algorithm where each node selfishly replicates objects from a distant server based on other nodes' placements and their reliability, can achieve good performance with minimal cost. Our results further evidence that the level of similarity of nodes' preferences within a social group is key for deciding which content placement strategy to follow within a replication group. The defined metrics of tightness which can be used as decision criteria when choosing content placement strategies under given group membership or, more broadly, for carrying out performance-driven group management operations such as group formation, merging or splitting.

The study of content dissemination in a networking environment comprised of interest and locality-induced social groups shows that when nodes exhibit the same preferences, mobility helps in the dissemination of content, and nodes attain an improvement even if they store content selfishly. Similarly, in a nomadic sensor network rational mobile nodes will exhibit cooperative behaviour on legs of the network, given that they actually transmit acquired sensor data with a probability p when they meet a selfish U-node. In a more realistic scenario, this content delivery probability would correspond to the reputation value of each U-node of being cooperative.

When disseminating or searching for content, we conclude that remarkably simple scaling behavior occurs with gossiping algorithms. Further, we discover tradeoffs between "smart" and "blind" gossiping schemes and algorithms. A smart gossiping algorithm where each node has full information about the informed or queried nodes in each round is, in nature, more effective than the blind selection scheme where no information is available. However, there is a trade-off between speed, cost and redundancy between these two schemes. A blind search process with smaller cost of managing information could compensate for the smaller speed in disseminating or finding a certain content, by querying more

nodes in each round. Further, a smart gossiping scheme actually burdens cooperative nodes with more redundant messages than a blind scheme, without producing an analogous gain in dissemination or search speed.

Based on the work conducted in this thesis, in future research we can proceed to study information dissemination and retrieval at a smaller level of detail, by taking into account the semantic and freshness characteristics of objects. Semantic characteristics are important for determining the relevance of an information object (owned by some node) with a certain data query, while freshness is important for determining the current value of an object, i.e., its depreciation due to its age. Ultimately, the goal is to design an information filtering algorithm that makes part of a dissemination or retrieval scheme, and assists in maximizing the benefit we attain from this procedure.

References

1. Jaho, E., Koukoutsidis, I., Stavrakakis, I., Jaho, I.: Cooperative replication in content networks with nodes under churn. In: Networking '08. (2008)
2. Jaho, E., Koukoutsidis, I., Stavrakakis, I., Jaho, I.: Cooperative content replication in networks with autonomous nodes. Computer Communications **35** (March 2012) 637–647
3. Jaho, E., Karaliopoulos, M., Stavrakakis, I.: Social similarity as a driver for selfish, cooperative and altruistic behavior. In: 4th IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC'10), Montreal, Canada (June 2010)
4. Jaho, E., Karaliopoulos, M., Stavrakakis, I.: Social similarity favors cooperation: the distributed content replication case. In: IEEE Transactions on Parallel and Distributed Systems, submitted. (April 2012)
5. Jaho, E., Karaliopoulos, M., Stavrakakis, I.: IsCoDe: a framework for interest similarity-based community detection in social networks. In: 3rd International Workshop on Network Science for Communication Networks, (NetSciCom'11), Shanghai, China (April 2011)
6. Jaho, E., Stavrakakis, I.: Joint interest- and locality-aware content dissemination in social networks. In: WONS'09: Proceedings of the Sixth international conference on Wireless On-Demand Network Systems and Services, Piscataway, NJ, USA, IEEE Press (2009) 161–168
7. Koukoutsidis, I., Jaho, E., Stavrakakis, I., Jaho, I.: Cooperative content retrieval in nomadic sensor networks. In: IEEE Infocom MOVE (Mobile Networking for Vehicular Environments) Workshop. (April 2008)
8. Tang, S., Jaho, E., Stavrakakis, I., Koukoutsidis, I., Mieghem, P.V.: Modeling gossip-based content dissemination and search in distributed networking. Comput. Commun. **34** (May 2011) 765–779
9. Cohen, E., Shenker, S.: Replication strategies in unstructured peer-to-peer networks. SIGCOMM Comput. Commun. Rev. **32** (August 2002) 177–190
10. La, C.A., Michiardi, P., Casetti, C., Chiasserini, C.F., Fiore, M.: A lightweight distributed solution to content replication in mobile networks. In: 2010 IEEE Wireless Communication and Networking Conference, IEEE (April 2010) 1–6
11. La, C.A., Michiardi, P., Casetti, C., Chiasserini, C.F., Fiore, M.: Content replication and placement in mobile networks. CoRR **abs/1102.3013** (2011)

12. Derhab, A., Badache, N.: Data replication protocols for mobile ad-hoc networks: A survey and taxonomy. Communications Surveys and Tutorials Journal **11** (2009)
13. Lim, S., Yu, C., Das, C.: Clustered mobility model for scale-free wireless networks. (November 2006) 231–238
14. Leff, A., Wolff, J., Yu, P.: Replication algorithms in a remote caching architecture. IEEE Trans. Par. and Distr. Systems **4**(11) (November 1993) 1185–1204
15. Laoutaris, N., Telelis, O., Zissimopoulos, V., Stavrakakis, I.: Distributed selfish replication. IEEE Trans. Par. Distr. Systems **17**(12) (December 2006) 1401–1413
16. Allen, S.M., Colombo, G., Whitaker, R.M.: Cooperation through self-similar social networks. ACM Trans. Auton. Adapt. Syst. **5** (February 2010) 4:1–4:29
17. Newman, M.E.J.: Detecting community structure in networks. The European Physical Journal B - Condensed Matter and Complex Systems **38**(2) (March 2004) 321–330
18. Newman, M.: Analysis of weighted networks. Phys. Rev. E **70**(5) (November 2004) 056131
19. Boldrini, C., Conti, M., Passarella, A.: Contentplace: Social-aware data dissemination in opportunistic networks. In: The 11-th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM'08). (2008)
20. Carreras, I., Chlamtac, I., Woesner, H., Zhang, H.: Nomadic sensor networks. In: Proc. EWSN. (2005) 166–175
21. Srinivasan, V., Nuggehalli, P., Chiasserini, C., Rao, R.: Cooperation in wireless ad hoc networks. In: INFOCOM. (2003)
22. Crowcroft, J., Gibbens, R., Kelly, F., Ostring, S.: Modelling incentives for collaboration in mobile ad hoc networks. In: WiOpt. (2003)
23. Seredynski, M., Bouvry, P., Klopotek, M.A.: Evolution of cooperation in ad hoc networks under game theoretic model. In: MobiWac. (2006) 126–130
24. Félegyházi, M., Hubaux, J.P., Buttyán, L.: Nash equilibria of packet forwarding strategies in wireless ad hoc networks. IEEE Trans. Mob. Comput. **5**(5) (2006) 463–476
25. Buttyán, L., Dóra, L., Félegyházi, M., Vajda, I.: Barter-based cooperation in delay-tolerant personal wireless networks. In: WoWMoM. (2007) 1–6
26. Pittel, B.: On spreading a rumor. SIAM J. Appl. Math. **47** (March 1987) 213–223
27. Nekovee, M., Moreno, Y., Bianconi, G., Marsili, M.: Theory of rumour spreading in complex social networks. Physica A: Statistical Mechanics and its Applications **374**(1) (January 2007) 457–470
28. Karp, R., Schindelhauer, C., Shenker, S., Vocking, B.: Randomized rumor spreading. In: Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Washington, DC, USA, IEEE Computer Society (2000) 565–
29. Kempe, D., Dobra, A., Gehrke, J.: Gossip-based computation of aggregate information. In: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science. FOCS '03, Washington, DC, USA, IEEE Computer Society (2003) 482–
30. Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)

Channel Coding Techniques with Emphasis on Convolutional and Turbo Codes

Alexandros Katsiotis *

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
akats@di.uoa.gr

Abstract. In this thesis, a family of low complexity convolutional codes is constructed, by modifying appropriately the trellis diagram of punctured convolutional codes. The goal is to improve performance at the expense of a reasonable low increase of the trellis complexity. Many new convolutional codes of various code rates and values of complexity are provided. In many cases, a small increase in complexity can lead to a great improvement of performance, compared to punctured convolutional codes. Furthermore, a method is presented for designing new flexible convolutional codes, by combining the techniques of path pruning and puncturing. The new codes can vary their rate, as well as the complexity of their trellis diagram, and hence the computational complexity of the decoding algorithm, leading to coding schemes that manage more efficiently the system resources, compared to classical variable rate convolutional codes. The complexity of the trellis diagram affects the computational complexity of every trellis based decoding algorithm, like BCJR and its variations. Thus, the possibility of applying the aforementioned results using recursive convolutional encoders, which are used as constituent encoders in turbo codes, is investigated. The goal is to construct efficient flexible turbo coding schemes. Simulation results indicate that in specific ranges of the signal to noise ratio, a great decrease in the computational complexity of the decoding procedure can even result to a decrease in the bit error rate.

1 Introduction

Traditionally, an (n, k, m) convolutional code $\mathcal{C}$ can be represented by a "conventional" semi-infinite trellis diagram. Although the trellis is semi-infinite, it consists, after a short initial segment, of concatenated copies of an 1-section structure called *trellis module*. The conventional *trellis module* consists of 2^m "initial states", and 2^m "final states". Each "initial state" is connected by a directed edge to 2^k "final states". Furthermore, each edge is labeled by an n-size binary tuple, which corresponds to the output of the encoder during the specific state transition. Thus, the conventional trellis module contains $n2^{m+k}$ edge bits,

* Dissertation Advisor: Nicholas Kalouptsidis, Professor

and this is called the *trellis complexity* of the module [1]. When a trellis module is used for maximum likelihood decoding of $\mathcal{C}$, the decoding complexity is proportional to the trellis complexity of the module. Particularly, the trellis complexity of the conventional trellis module increases exponentially with k and m, which means that especially for high rate codes, maximum likelihood decoding becomes a very complex procedure.

In order to overcome this problem, punctured convolutional codes (PCC) were introduced by Cain et al. [2]. The trellis module of an (n, k, m) PCC is constructed by "blocking" k "conventional" trellis modules of an $(N, 1, m)$ "mother" code, and then deleting (puncturing) all but n edge bits from each output tuple of this "block". The semi-infinite trellis diagram of the PCC consists (after a short initial transient) of repeated copies of the k-section *punctured trellis module*. The trellis complexity of an (n, k, m) PCC is equal to $n2^{m+1}$ and it is significantly less than the complexity of the "conventional" trellis module of an (n, k, m) convolutional code. PCCs have been extensively studied for various rates [2–4]. Bocharova et al. [4] published comprehensive tables with the best (n, k) PCCs for $3 \leq n \leq 8$, $2 \leq k \leq n-1$.

Apart from the low decoding complexity, another significant advantage of PCCs is *rate variability*. That is, a whole family of PCCs of various code rates can be constructed, using a single mother code and a variety of puncturing matrices [3, 5]. Each one of the respective codes can be decoded using the decoder (and hence the trellis) of the mother code. Hence, depending on the channel conditions the suitable code is used, a fact that results in a more efficient use of bandwidth. The flexibility offered by PCCs is valuable for modern communication systems, like for instance the IEEE 802.22 standard [6]. Rate variability can be also achieved by using the technique of *path pruning* [7]. That is, using determinate and indeterminate information bits in order to prune away some codeword paths from the trellis of a convolutional code.

An (n, k, m) convolutional code C can be represented by various trellis diagrams (periodic in general), and hence various trellis modules [1, 8]. Apart from the trellis complexity, an important quantity associated with a trellis module is the total number of merges contained in it. The number of merges at a specific state is equal to the total number of branches reaching it minus one [9]. The trellis complexity and the total number of merges are equal to the number of real additions and real comparisons respectively, the Viterbi algorithm has to perform per trellis module [1, 9]. Hence, they constitute the computational complexity of the decoding procedure of C, for a specific trellis module. McEliece et al. demonstrate in [1] that any (n, k, m) convolutional code has a minimal trellis representation, which in most cases is significantly less complex (in terms of trellis complexity), than the conventional trellis diagram. A minimal trellis has the minimum possible complexity and can be constructed directly by a trellis canonical generator matrix. The minimal trellis module minimizes many other quantities, like the total number of states, the total number of merges, etc. They also show that PCCs are simply convolutional codes whose generator matrices have a special structure, and this fact explains the reason why they can be

represented by trellis diagrams of low complexity. Finally, they set under consideration the existence of other classes of low complexity convolutional codes, that contain good codes.

2 New Constructions of Low Complexity Convolutional Codes

In this section, we introduce a class of low complexity convolutional codes, produced by modifying the time varying punctured trellis module [10]. Our goal is to achieve better performance than the initial PCC code, at the expense of a reasonable low increase of the trellis complexity. The increase of the complexity is achieved by allowing variations in the state dimension of the trellis module.

We consider a periodic time-varying linear encoder of period k [11], described by the $1 \times n_t$ binary matrices $\mathbf{G}_j(t)$, for $0 \leq j \leq m_t$, such that

$$\boldsymbol{v}_t = \sum_{j=0}^{m_t} u_{t-j}\mathbf{G}_j(t), \qquad t \geq 0 \tag{1}$$

where u_t and $\boldsymbol{v}_t$ denote the input value of size 1 and output value of size n_t respectively at time instant t. By way of convention, $\mathbf{G}_j(t) = \mathbf{0}$ for $j < 0$ and $j > m_t$, where m_t is the memory of the encoder at the t-th time instant. It holds, $\mathbf{G}_j(t+k) = \mathbf{G}_j(t)$, $m_{t+k} = m_t$, and $n_{t+k} = n_t$ for all j and t.

The trellis diagram representing the above construction is periodic (after a short initial transient) and consists of repeated copies of a k-section structure, called *trellis module*. During a trellis module, k information bits are associated to n encoded bits, where $n = \sum_{t=0}^{k-1} n_t$. The state space dimension of the trellis module at the t-th time instant is m_t, $0 \leq t \leq k-1$. We call the time interval of the trellis module between the t-th and the $(t+1)$-th time instant, as the t-th *trellis section*. $v_{t,i}$ denotes the i-th output bit which is transmitted during the t-th trellis section, for $1 \leq i \leq n_t$, and for $0 \leq t \leq k-1$, and is produced by the generator sequence $\mathbf{g}^{(t,i)} = [g_0^{(t,i)}\ g_1^{(t,i)} \cdots g_{m_t}^{(t,i)}]$.

Given a positive integer m, we impose the following restrictions: $m_t \in \{m, m+1\}$, and $m_0 = m$. We denote by T_{ζ_1,ζ_2} the set of *trellis sections* t, $0 \leq t \leq k-1$, during which the state dimension changes from ζ_1 to ζ_2, where $\zeta_1, \zeta_2 \in \{m, m+1\}$. That is, $T_{\zeta_1,\zeta_2} = \{t | m_t = \zeta_1,\ m_{t+1} = \zeta_2,\ 0 \leq t \leq k-1\}$. We further confine the code structure so that the encoding matrices $\mathbf{G}_j(t) = [g_j^{(t,1)}\ g_j^{(t,2)} \cdots g_j^{(t,n_t)}]$ satisfy the following properties:

P_1. $\mathbf{G}_0(t) = \mathbf{1}$, $0 \leq t \leq k-1$,
P_2. $\mathbf{G}_m(t) = \mathbf{1}$, when $t \in T_{m,m}$,
P_3. $\mathbf{G}_{m+1}(t) = \mathbf{1}$, when $t \in T_{m+1,m+1}$,
P_4. $g_m^{(t,i)} \neq 0$ or $g_{m+1}^{(t,i)} \neq 0$, for all $1 \leq i \leq n_t$, and $\mathbf{G}_m(t)$ and $\mathbf{G}_{m+1}(t)$ are linearly independent, when $t \in T_{m+1,m}$.

Note that from Property P_4, it holds that $n_t \geq 2$, when $t \in T_{m+1,m}$. We denote by p_t the smallest integer, $1 \leq p_t < n_t$, such that

$$g_m^{(t,p_t+1)} = \cdots = g_m^{(t,n_t)} \quad \text{and} \quad g_{m+1}^{(t,p_t+1)} = \cdots = g_{m+1}^{(t,n_t)}$$

when $t \in T_{m+1,m}$.

The *trellis complexity* (TC) of a T-section *trellis module* M is defined as the total number of edge symbols contained in the module per information bit [1]. That is, $TC(M) = \sum_{t=0}^{T-1} n_t 2^{m_t+k_t}$, where k_t is the number of information bits associated with the t-th trellis section. It is straightforward to verify that the trellis complexity of the new module M associated with the new code is given by

$$TC(M) = \sum_{t \in T_{m,m}} n_t 2^{m+1} + \sum_{t \in T_{m,m+1}} n_t 2^{m+1} + \sum_{t \in T_{m+1,m+1}} n_t 2^{m+2} + \sum_{t \in T_{m+1,m}} n_t 2^{m+2} \quad (2)$$

since $k_t = 1$, for $0 \leq t \leq k-1$.

Theorem 1. *The complexity of the minimal trellis module M_{min} of $\mathcal{C}$ is given by*

$$TC(M_{min}) = \sum_{t \in T_{m,m}} n_t 2^{m+1} + \sum_{t \in T_{m,m+1}} n_t 2^{m+1} + \sum_{t \in T_{m+1,m+1}} n_t 2^{m+2} + \sum_{t \in T_{m+1,m}} (n_t + p_t) 2^{m+1}. \quad (3)$$

Comparison of (3) and (2), indicates that the initial trellis module is not minimal. Next, we show that if we replace every trellis section $t \in T_{m+1,m}$ by an equivalent pair of trellis sections, the resulting trellis module is minimal.

Theorem 2. *The trellis section $t \in T_{m+1,m}$ is equivalent to a pair of trellis sections, where the first trellis section t_1 carries the first p_t of the n_t output bits and has one bit input. Furthermore, it has 2^{m+1} initial states and 2^{m+1} final states. The second section t_2 produces the last $n_t - p_t$ output bits of the original trellis section t and it is informationless. Moreover, it has 2^{m+1} initial states and 2^m final states.*

The first and second sections of the pair contain $p_t 2^{m+2}$ and $(n_t - p_t)2^{m+1}$ edge symbols respectively (the second section is informationless, thus only one edge diverges from each initial state). Thus, in total the pair of sections contains $(n_t + p_t)2^{m+1}$ edge symbols. If we replace in (2) the term $\sum_{t \in T_{m+1,m}} n_t 2^{m+2}$ with $\sum_{t \in T_{m+1,m}} (n_t + p_t) 2^{m+1}$, relation (3) results. That is, by simply replacing in the initial trellis module of a code $\mathcal{C}$ each one of the trellis sections $t \in T_{m+1,m}$ with its equivalent pair of sections, the result is a minimal trellis module for $\mathcal{C}$.

2.1 Search Results

The proposed search procedure initializes with the punctured trellis module of an (n, k, m) convolutional code, and increases the state space dimension of specific

trellis sections from m to $m+1$, performing search over the generator sequences $\mathbf{g}^{(t,i)}$. The objective is to determine codes of better spectra at the expense of higher trellis complexity. However complexity is maintained at affordable levels, as it remains less than the trellis complexity of the respective $(n, k, m+1)$ PCC. The search was conducted over a range of rates and memory sizes, and many good codes were found. The following comments are worth to point out:

- Some of the new codes achieve greater d_f than the respective PCCs of memory m, at the expense of a small increase of the TC.
- Some of the new codes achieve d_f equal to the d_f of the PCCs of memory $m+1$. That is, they have similar asymptotic performance (for high signal to noise ratio) with the corresponding $(n, k, m+1)$ PCCs but less trellis complexity.
- Many of the new codes achieve a specific d_f, with the least known trellis complexity (for specific rate).
- In many other cases, a small increase of TC was enough to significantly improve the spectra (for equal d_f) of the best PCC of memory m.

3 Flexible Convolutional Codes: Variable Rate and Complexity

In this section, we present a method that combines the techniques of path-pruning and puncturing, in order to construct convolutional codes that can vary both their rate and the computational complexity of the decoding procedure [12]. Starting from an $(n, 1, m)$ mother convolutional code, we construct a large family of codes of various code rates. For each code rate the family contains trellis modules of various values of computational complexity. Path pruning is used in order to remove from the mother trellis an amount of state transitions (branches), a fact that results into a less complex trellis. Furthermore, puncturing is utilized for adjusting the code rate. Particularly, the two aforementioned techniques are employed as follows.

Let u_t be the information bits and $\hat{u}_t$ the input bits of the mother encoder. The path-pruning on the trellis of a convolutional code can be implemented by mapping the information u_t and register state bits into the final input $\hat{u}_t$ of the encoder.

More precisely, every T_{pr} time units the single input bit of the encoder is not an information bit, rather it is computed as a linear combination of bits of the current state $S_t = \{\hat{u}_{t-1}, \cdots, \hat{u}_{t-m}\}$, i.e.

$$\hat{u}_{t_1 T_{pr}+t_2} = \begin{cases} u_{t_1(T_{pr}-1)+t_2} & , \text{ if } t_2 \neq 0 \\ \sum_{i=1}^{\hat{d}} c_i \hat{u}_{t_1 T_{pr}-i}, & \text{ if } t_2 = 0 \end{cases} \tag{4}$$

where $t_1 = \left\lfloor \frac{t}{T_{pr}} \right\rfloor$, $t_2 = t \bmod T_{pr}$, $t = 1, 2, \ldots$, and $\hat{d}$ is the degree of the polynomial $c(X) = \sum_{i=1}^{m} c_i X^i$. The binary coefficients are chosen, such that

$c_i = 0$ for $0 = i \bmod T_{pr}$, i.e. the non information bits are produced by the linear combination of information bits only.

Path-pruning results to an $(nT_{pr}, T_{pr} - 1)$ time-varying convolutional code. The resulting semi-infinite trellis diagram starts at the all zero state, and after a short initial transient, it becomes periodic, consisting of concatenations of a T_{pr}-section trellis module M_{pr}.

Theorem 3. *The state space dimension s_l^{pr} at depth l and the branch space dimension b_l^{pr} of the l-th trellis section of trellis module $\mathcal{M}_{pr}$ are given by*

$$s_l^{pr} = \begin{cases} m - 1 - \hat{\beta}, & \text{for } 1 \le l \le \hat{\alpha} \\ m - \hat{\beta}, & \text{for } \hat{\alpha} + 1 \le l \le T_{pr} \end{cases}$$

$$b_l^{pr} = \begin{cases} m - \hat{\beta}, & \text{for } 1 \le l \le \hat{\alpha} \text{ or } l = T_{pr} \\ m + 1 - \hat{\beta}, & \hat{\alpha} + 1 \le l \le T_{pr} - 1 \end{cases}$$

where $\hat{\alpha} = m - \hat{d} \bmod T_{pr}$, $\hat{\beta} = \left\lfloor \frac{m - \hat{d}}{T_{pr}} \right\rfloor$.

The final trellis module M_{pu} is constructed by concatenating p copies of the M_{pr} module, and puncturing a portion of the encoded bits based on a specific puncturing matrix P, in order to adjust the rate. The final trellis module consists of $T_{pu} = pT_{pr}$ trellis sections. The total number of merges E_{pu} contained in M_{pu} are

$$E_{pu} = p(T_{pr} - \frac{\hat{a}}{2} - 1) \cdot 2^{m - \hat{b}}. \tag{5}$$

Depending on the puncturing matrix, the trellis complexity of the final trellis module M_{pu} is equal to

$$TC(M_{pu}) = \sum_{j=1}^{pT_{pr}} n_j 2^{b_{j \bmod T_{pr}}^{pr}}, \tag{6}$$

where $0 \le n_j \le n$ is the number of output bits of the j-th section of M_{pu}, for $1 \le j \le pT_{pr}$. The final trellis module M_{pu} corresponds to a $(pT_{pr}n - \tilde{n}, p(T_{pr} - 1))$ convolutional code, where $\tilde{n}$ is the number of the punctured encoded bits, i.e. $pT_{pr}n - \tilde{n} = \sum_{j=1}^{pT_{pr}} n_j$.

Given an $(n, 1, m)$ mother code, for various choices of T_{pr}, p, $c(X)$ and puncturing matrix P, arbitrarily many codes of different code rates and values of computational complexity can be constructed. In this study, we restrict ourselves to a subset according to the following criterion.

Given a rate k'/n', we construct codes with computational complexity (i.e. trellis complexity and number of merges) equal to the computational complexity of the trellis module of a PCC with memory size m', for various values of m'.

3.1 Results and Simulations

In this study we present three families of flexible codes, generated by a $(2, 1, 8)$, a $(3, 1, 8)$ and a $(4, 1, 8)$ mother convolutional code respectively. All codes are of

rates $(n'-1)/n'$, for $2 \leq n' \leq 8$. Many of the codes have the same free distance with the best codes of the same rate and complexity presented in [4, 13].

We have simulated the performance of various codes produced by the $(4, 1, 8)$ mother convolutional code, for the AWGN channel, using BPSK modulation. Some of the results are depicted in Figure 1, which indicates BER plots for all values of m' and rates 1/2, 2/3 and 3/4.

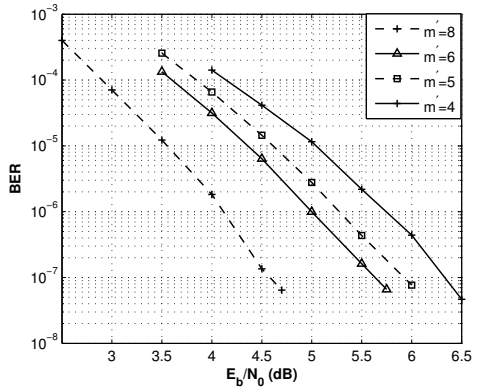
In contrast to the classical variable rate convolutional codes [3, 5] (i.e. a mother convolutional encoder and a set of puncturing matrices), which "exchange" coding gain for bandwidth and vice versa, the proposed constructions add one more "dimension", i.e. the computational complexity of decoding, hence leading to coding schemes that manage more efficiently the system resources. Indeed, consider the family of codes produced by the $(4, 1, 8)$ mother convolutional code. Assume that the SNR is equal to 7.1dB, and that the code of rate 3/4 and $m' = 4$ ($TC = 42.7$, $E_{pu} = 16$, all normalized by k') is used. Then, an error probability of 10^{-7} is achieved (Fig. 1(c)). Furthermore, assume that the SNR decreases by 0.8dB. There are many strategies that can be followed during the next transmission, for the anticipation of the SNR's decrease. For instance, one choice is the use of the code of rate 1/2 and $m' = 4$ ($TC = 64$, $E_{pu} = 16$) (Fig. 1(a)), i.e. decreasing the rate and keeping the complexity in similar levels. Another option is the use of the code of rate 3/4 and $m' = 6$ (Fig. 1(c)) ($TC = 170.7$, $E_{pu} = 64$), i.e. keeping the rate constant and increasing the complexity. Finally, the code of rate 2/3 and $m' = 5$ ($TC = 96$, $E_{pu} = 32$) (Fig. 1(b)) could be used, resulting in a small decrease in the rate and a small increase in complexity.

4 Recursive Flexible Convolutional Encoders for Parallel Concatenation

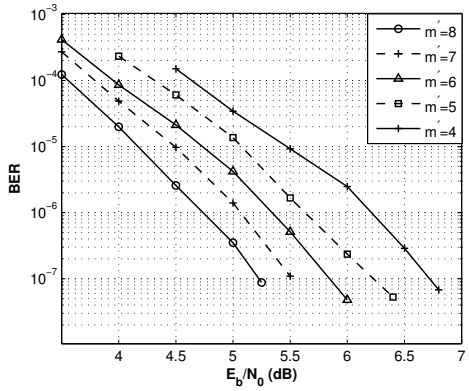
In this section, we extend the previous analysis and we combine path-pruning and puncturing for the construction of flexible convolutional codes for parallel concatenation that can vary both the rate and the computational complexity of the decoding procedure, leading to flexible turbo codes.

Consider a trellis module as it is constructed in section 3, that corresponds to a code of rate k'/n' and has the complexity profile of the respective PCC with memory m'. It can be shown that if the Max-Log-MAP algorithm is used for decoding, then $(n'+q+2k')2^{m'+1}+k'$ additions and $4k'2^{m'}-k'$ comparisons have to be performed in order to decode k' information bits. q denotes the number of trellis section that carry encoded bits

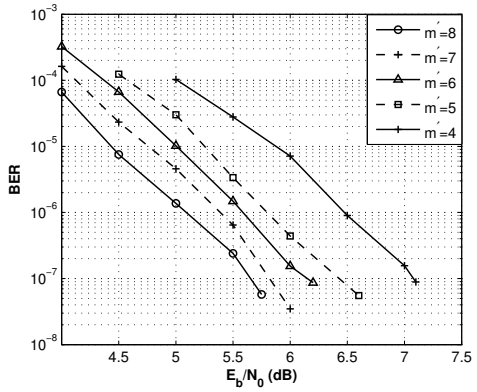
Consider a feedforward $(n, 1, m)$ mother convolutional encoder, described by the polynomial generator matrix $G(D) = \left[g^{(0)}(D) \quad g^{(1)}(D) \ldots g^{(n-1)}(D)\right]$. Assume that for particular path-pruning and puncturing parameters (i.e. $c(X)$, T_{pr}, p, P), a specific code C_{pu} and its respective trellis module M_{pu} are constructed. Consider also the equivalent $(n, 1, m)$ recursive systematic mother encoder given by $G_{sys}(D) = \left[1 \quad \frac{g^{(1)}(D)}{g^{(0)}(D)} \ldots \frac{g^{(n-1)}(D)}{g^{(0)}(D)}\right]$. It can be shown that by using the same



(a)



(b)



(c)

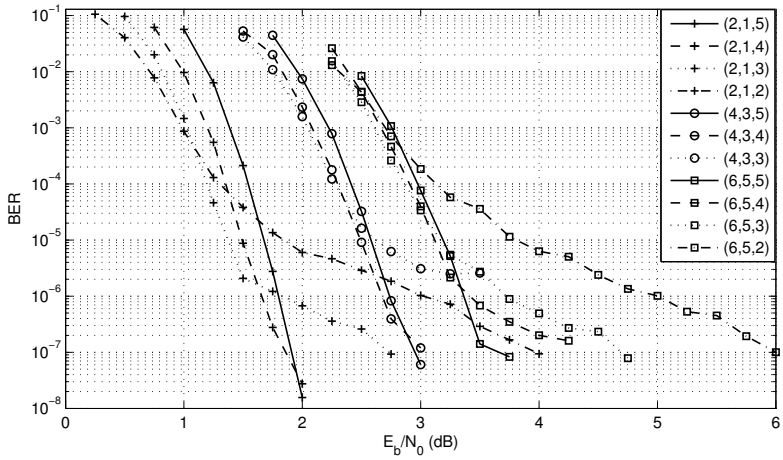
Fig. 1. BER simulations of codes produced by the $(4, 1, 8)$ mother convolutional code, for various values of m' and rates (a) 1/2, (b) 2/3 and (c) 3/4.

parameters T_{pr}, p, P and the pruning polynomial $c'(X) = \sum_{i=1}^{m}(c_i + g_i^{(0)})X^i$ instead of the polynomial $c(X)$, the same code C_{pu} is produced and a corresponding trellis module M'_{pu}. c_i and $g_i^{(0)}$ are the coefficients of polynomials $c(X)$ and $g^{(0)}(D)$ respectively. Trellis modules M_{pu} and M'_{pu} have identical complexity profiles and they produce the same code. However, they correspond to different encoders. Hence, we can apply the results and the analysis presented in section 3, in case where recursive mother convolutional encoders are used.

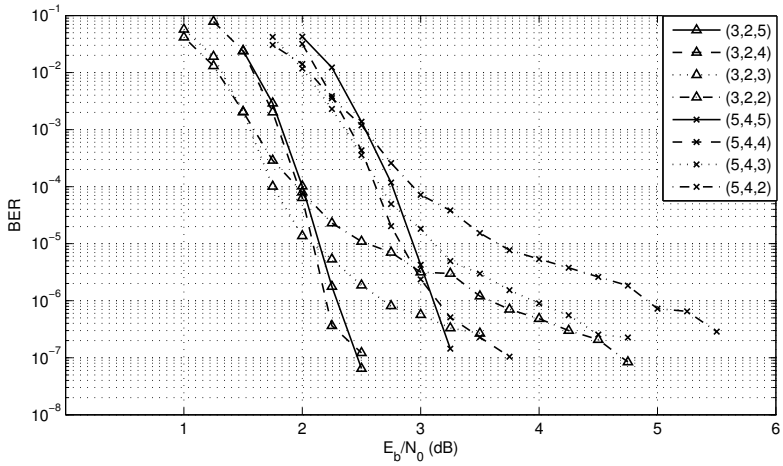
In this thesis we provide a family of recursive systematic convolutional encoders constructed from a $(2, 1, 5)$ recursive systematic mother encoder. All encoders are of rates $(n' - 1)/n'$, for $2 \leq n' \leq 6$. In [14], the authors provide the best recursive systematic punctured convolutional encoders for parallel concatenation, constructed from $(2, 1)$ mother encoders. Variable rate encoders are also provided. Almost all the encoders constructed in this thesis achieve the same value of distance d_2, as the encoders in [14].

We have simulated the performance of turbo codes that use the flexible mother encoder presented in this thesis and its associative family, as constituent encoders, for the AWGN channel using BPSK modulation and the Max-Log-MAP decoding algorithm. In Fig. 2 we have used a random interleaver of length 1000 for all the members of the family. Furthermore, for a particular encoder and a particular SNR value we have assumed a fixed number of iterations L_{SNR}^f, which is less than or equal to 10. Particularly, L_{SNR}^f is equal to the minimum number of iterations needed to achieve the minimum possible error probability, if the former is less than 10, and equal to 10 otherwise. Note that, apart from the computational complexity of the decoding algorithm that uses a particular trellis module, the overall computation complexity of turbo decoding depends on the number of iterations as well. Thus, in Table 1 we provide the values of L_{SNR}^f associated with the BER plots in Fig. 2. Particularly, the first row of Table 1 indicates the curve points. A specific curve point for a specific encoder corresponds to a particular value of SNR (Fig. 2). As depicted in Fig. 2, the *waterfall region* for the (n', k', m') constituent codes (of all rates) for $m' < m$ appears at smaller values of SNR than the waterfall region in case where the respective (n', k', m) codes of the family (including the mother code) are used. This comes with a significant reduction in computational complexity. In other words, in particular SNR regions, a reduction in complexity can result in an increase of performance. As an example, the $(2, 1, 5)$ mother code achieves error probability $2 \cdot 10^{-4}$ for SNR equal to 1.5dB (Fig. 2a). The $(2, 1, 3)$ code, for the same SNR value, achieves error probability $2 \cdot 10^{-6}$ and reduces the computational complexity by 75%. Similar behavior can be observed for all code rates. As a matter of fact, there is no need to use the (most complex) (n', k', m) code of each rate for error probabilities greater than 10^{-6}.

The proposed constructions can lead to turbo coding schemes that manage more efficiently the system resources. Assume that the SNR is equal to 1.75dB, and that the turbo code uses the $(2, 1, 4)$ (9 iterations) convolutional encoder from the family, achieving error probability $3 \cdot 10^{-7}$ (Fig. 2a). Furthermore, assume that the SNR increases by 1.75dB. In this case, there is no need for using a



(a)



(b)

Fig. 2. Simulations of turbo codes that use as constituent encoders the constructed family of flexible convolutional encoders.

powerful constituent encoder. During the next transmission we can increase the rate, or decrease the complexity. For instance, we can use the $(5, 4, 4)$ constituent encoder (Fig. 2b) (7 iterations), increasing significantly the rate and keeping the same value of m'. Furthermore, the $(2, 1, 2)$ encoder (Fig. 2a) (3 iterations) can be used, i.e. keeping the rate constant and decreasing substantially the complexity. At this point, one may wonder whether the reduction in computational complexity can be achieved by simply reducing the number of iterations. The computational complexity that corresponds to 3 iterations of the $(2, 1, 2)$ decoder is slightly less than the complexity of one iteration of the $(2, 1, 4)$ decoder which was initially used. The $(2, 1, 4)$ code achieves error probability $6 \cdot 10^{-5}$ at SNR equal to 3.5dB, if only one iteration is performed.

Table 1. number of iterations of the codes in fig. 2

code	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
(2,1,5)	3	8	10	10	10											
(2,1,4)	5	7	10	10	9	9										
(2,1,3)	4	7	10	10	10	9	7	6	5	3						
(2,1,2)	3	6	10	10	10	8	8	7	5	5	5	3	3	3	3	3
(3,2,5)	7	10	10	10	9											
(3,2,4)	2	5	10	10	10	10										
(3,2,3)	5	6	10	10	10	8	8	7	4	3	3					
(3,2,2)	5	7	10	10	8	8	7	7	6	4	4	4	3	3	3	3
(4,3,5)	5	10	10	10	10	10										
(4,3,4)	5	6	10	10	10	10	8									
(4,3,3)	4	8	10	10	10	10	7	6	5							
(5,4,5)	3	6	10	10	10	10										
(5,4,4)	4	9	10	10	10	10	7	6								
(5,4,3)	3	7	10	10	9	8	8	7	6	5	5	4	4			
(5,4,2)	3	5	7	8	8	8	7	6	6	6	5	3	3	3	2	2
(6,5,5)	7	10	10	10	10	7										
(6,5,4)	5	10	10	10	10	7	6	6	6							
(6,5,3)	5	9	10	10	9	8	7	6	6	5	5					
(6,5,2)	4	6	10	10	8	7	7	6	5	5	3	3	3	2	2	2

5 Conclusion

In this thesis, a family of low complexity convolutional codes was proposed, which was constructed by modifying appropriately the trellis diagram of punctured convolutional codes. Many new convolutional codes of various code rates and values of complexity were provided. Furthermore, a method was presented for designing new flexible convolutional codes that can vary both their rate and the computational complexity of the decoding procedure, leading to coding schemes that manage more efficiently the system resources, compared to classical variable rate

convolutional codes. This method was extended in order to construct recursive flexible convolutional encoders. Their use as constituent convolutional encoders in turbo codes can lead to pretty flexible parallel concatenated coding schemes.

References

1. R. J. McEliece and W. Lin, "The trellis complexity of convolutional codes," *IEEE Trans. Inf. Theory*, vol. IT-42, pp. 1855–1864, Nov. 1996.
2. J. B. Cain, J. C. Clark, Jr, and J. M. Geist, "Punctured convolutional codes of rate $(n-1)/n$ and simplified maximum likelihood decoding," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 97–100, Jan. 1979.
3. P. J. Lee, "Constructions of rate $(n-1)/n$ punctured convolutional codes with minimal required SNR criterion," *IEEE Trans. Commun.*, vol. COM-36, pp. 1171–1173, Oct. 1988.
4. I. E. Bocharova and B. D. Kudryashov, "Rational rate punctured convolutional codes for soft-decision Viterbi decoding," *IEEE Trans. Inf. Theory*, vol. IT-43, pp. 1305–1313, July 1997.
5. G. Begin, D. Haccoun, and C. Paquin, "Further results on high-rate punctured convolutional codes for Viterbi and sequential decoding," *IEEE Trans. Commun.*, vol. COM-38, pp. 1922–1928, Nov. 1990.
6. C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. Shellhamer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Commun. Mag.*, vol. 47, pp. 130–138, Jan. 2009.
7. C. H. Wang and C. C. Chao, "Path-compatible pruned convolutional (PCPC) codes," *IEEE Trans. Commun.*, vol. COM-50, pp. 213–224, Feb. 2002.
8. B. F. Uchoa-Filho, R. D. Souza, C. Pimentel, and M. Jar, "Convolutional codes under a minimal trellis complexity measure," *IEEE Trans. Commun.*, vol. COM-57, pp. 1–5, Jan. 2009.
9. R. J. McEliece, "On the BCJR trellis for linear block codes," *IEEE Trans. Inf. Theory*, vol. IT-42, pp. 1072–1092, July 1996.
10. A. Katsiotis, P. Rizomiliotis, and N. Kalouptsidis, "New constructions of high performance low complexity convolutional codes," *IEEE Trans. Commun.*, vol. 58, pp. 1950–1961, July 2010.
11. M. Mooser, "Some periodic convolutional codes better than any fixed code," *IEEE Trans. Inf. Theory*, vol. IT-29, pp. 750–751, Sept. 1983.
12. A. Katsiotis, P. Rizomiliotis, and N. Kalouptsidis, "Flexible convolutional codes: variable rate and complexity," *IEEE Trans. Commun.*, March 2012.
13. E. Rosnes and Ø. Ytrehus, "On maximum length convolutional codes under a trellis complexity constraint," *Journal of Complexity*, vol. 20, pp. 372-408, March-June 2004.
14. F. Daneshgaran, M. Laddomada, and M. Mondin, "High-rate recursive convolutional codes for concatenated channel codes," *IEEE Trans. Commun.*, vol. COM-52, pp. 1846–1850, Nov. 2004.

MENTOR: An Emotional Tutoring Model for Distance Learning
MENTOR's Application in the Field of Didactics of Informatics

Makis Leontidis

Department of Informatics and Telecommunications, University of Athens
Panepistimiopolis, GR-15784 Athens, Greece
leon@di.uoa.gr·

Abstract. The aim of this dissertation is to present the MENTOR (Modelling EmotioNal TutORing) which is an emotional learning model that uses an Affective Module in order to recognize the affective state of the student during his interaction with the educational environment and thereafter to provide him with a suitable learning strategy constructing in this way an affective learning path. MENTOR constitutes of three main components, the Emotional Component, the Teacher Component and the Visualisation Component and its main purpose is to motivate appropriately the student in order to accomplish his learning goals. The basic concern of MENTOR is to retain the student's emotional state positive during the learning process. To achieve this, it recognises the emotions of the students and takes them under consideration to provide them with the suitable learning strategy. This kind of strategy is based both on the cognitive abilities and the affective preferences of the student and is stored in the student's model. The student model supplies the educational system with necessary information with the aim to adapt itself successfully to the student's needs.

1 Introduction

During the learning process in the real class, a creative teacher usually invests a significant amount of his efforts and time to identify the personality and the mood of his students in order to find the suitable ways of increasing their motivation [1]. The intrinsic ability of a good teacher to balance subtly and accurately their students' emotional predisposition, their individual needs and preferences and their current disposition, while he directs them adroitly to their goals' achievement, is one of the major factors to the student's progress and successful attainment of learning. Despite the importance of the affective factor, in most educational systems, this crucial parameter seems to have been ignored, since the significant process of learning is supported by methods which are mainly concentrating on the cognitive abilities of the student.

As a result, few contemporary educational systems began to consider their operation under an affective perspective with the aim of modelling the emotional processes which are taking place during the educational session [2]. Corresponding affective techniques are being incorporated more frequently in educational systems

· Dissertation Advisor Constantin Halatsis, Professor

with the aim of recognising student's emotions, mood and personality [3]. The traditional student model starts to be modified in order to be capable of storing affective information.

According to this point of view, we developed the MENTOR which is an Affective Educational Module capable of supporting the learning in the distance education [4]. MENTOR consists of three main components, which are the Emotional, the Teacher and the Visualisation Components respectively. MENTOR takes into account the personality and the emotional state of the student, in order to decide which is the appropriate affective tactic for him. Taking the above points into consideration, it seems clearly that the main purpose of the MENTOR is to create or to maintain a positive mood to the student, keeping him in track of his learning goals. To achieve this, we need to be aware of the student's emotional state in every moment. That is stored in the affective student model, which consists of cognitive and emotional information, and it is provided by the Emotional Component. In accordance with this plan, the model selects and supplies accurately the student with the proper affective tactics. In this manner, it involves effectively the student into the learning process under a fruitful pedagogical perspective.

2 Basic Issues of Affective Computing

The term Affective Computing involves the intention of Artificial Intelligence researchers to model emotions in intelligent systems. According to Picard [5] an affective system must be capable of recognizing emotions, respond to them and react "emotionally". Among the basic terms that determine the affective computing are personality, mood and emotions.

The personality determines all those characteristics that distinguish one human being from another. It is related to its behaviour and mental processes and has a permanent character [6]. The most known model of personality is the Five Factor Model (FFM) and results from the study of Costa and McCrae [7]. It is a descriptive model with five dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Due to these dimensions the model is also called OCEAN model. The descriptive character of FFM and the particular characteristics that accompany each type of personality (traits) allow us to model the student's personality [3] and use this information in educational applications [2]. The FFM provides us with a reliable way in order to connect a student's personality with his mood and emotions that he possibly develops during the learning process. This is very useful because we are able to initiate a student's emotional state and select the suitable pedagogical strategy.

Mood is a prolonged state of mind, resulting from a cumulative effect of emotions [6]. Mood differs from the emotion because it has lower intensity and longer duration. It can be consequently considered that mood is an emotional situation more stable than emotions and more volatile than personality. Based on this definition we categorize mood into two categories named, positive and negative. We consider that the student has either a positive mood when he feels emotions like joy, pride, hope, satisfaction, gratification, love, or a negative mood when feels emotions like sadness,

fear, shame, frustration, anger, disappointment, anxiety. Depending on this mood we speculate the possible emotions of the student.

The emotion is the synchronized response for all or most organic systems to the evaluation of an external or internal event. Emotion is analogous to a state of mind that is only momentary. Nevertheless, various attempts have been made, but the cognitive theory of emotions, known as OCC model, which formulated by Ortony, Clore and Collins [8], keeps a distinctive position among them. The three authors constructed a cognitive theory of emotion that explains the origins of emotions, describing the cognitive processes that elicit them. The OCC model provides a classification scheme for 22 emotions based on a valence reaction to events, objects and agents.

In our work we adopt the OCC model, because it elicits the origin of emotions under a cognitive aspect and it is possible to be computerized. So, based on this model we are able to classify and interpret a student's emotions in the learning process. The authors of the OCC model consider that it could be computationally implemented and help us to understand which are the emotions that the human beings feel, and under which conditions. Furthermore, they believe that relying on this model we could predict and explain human reactions to the events and objects. This is the main reason we use the OCC model in our study. The perspective by which, we construct the following component is interdisciplinary and focuses in the intersection of Artificial Intelligence and Cognitive Psychology.

3 The architecture of the MENTOR

MENTOR is an "affective" module which aims to recognize the emotions of the student during his interaction within an educational environment and thereafter to provide him with a suitable learning strategy [9]. The operation of MENTOR is based on the FFM [7] and the OCC model [8]. The module is being attached to an Educational System providing the system with the essential "emotional" information in order to determine the strategy of learning in collaboration with the cognitive information. The architecture of MENTOR is presented in Figure 1.

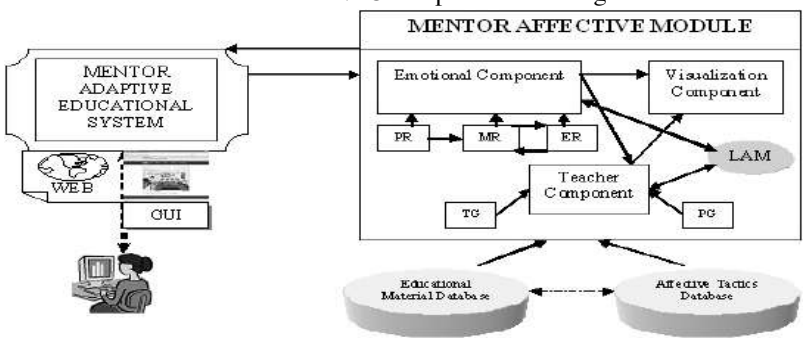


Fig. 1. The architecture of the Mentor

The MENTOR has three main components: The Emotional Component (EC), the Teacher Component (TC) and the Visualization Component (VC), which are respectively responsible for: a) the recognition of student's personality, mood and

emotions during the learning process, b) the selection of the suitable teaching and pedagogical strategy and c) the appropriate visualization of the educational environment. The combined function of these components "feeds" the AES with the affective dimension optimizing the effectiveness of the learning process and enhancing the personalized teaching.

The architecture of the MENTOR is designed with equal respect to the cognitive and the emotional dimension of teaching as well. So, we consider that the Teacher Component which is in charge of the formation of teaching consists of two sub-components, the Teaching Generator and the Pedagogical Generator which are responsible for providing the cognitive and emotional tactic respectively. Therefore, we use the term affective tactic so as to denote that the learning method which is suggested by the Teacher Component is a two-dimensional combination of cognitive and emotional guidance and support. The main purpose of MENTOR is to create the appropriate learning environment for the student, taking into account particular affective factors in combination with cognitive abilities of the student offering in this way personalized learning.

3.1 Recognizing the emotions of the student

The necessity of recognizing the student's emotion during the learning process, especially in distant learning environments is crucial and has been pointed out by many researchers in the e-learning field ([10],[11]). Concerning the MENTOR, responsible for the recognition of the student's emotions is the Emotional Component. This component (Figure 1) is composed by three subcomponents, the Personality Recognizer (PR), the Mood Recognizer (MR) and the Emotion Recognizer (ER), which are responsible for the recognition of the personality, mood and emotions of the student. When the student uses the system for the first time, the PR subcomponent assesses the type of a student's personality. As a result, the student's traits are being recognized and are being used by the Teacher Component for the suitable selection of pedagogical and teaching strategy. For example, a student that has been recognized as Openess, according to FFM is imaginative, creative, explorative and aesthetic [7]. These characteristics are evaluated by the TC providing the system with an exploratory learning strategy, giving more autonomy of learning to the student and limiting the guidance of the teacher. The MR subcomponent recognizes and categorizes the student's mood either as positive or as negative. In our approach, good mood consists of emotions like joy, satisfaction, pride, hope, gratification and bad mood consists of emotions like distress, disappointment, shame, fear, reproach. As a result, we have an initial evaluation of the current emotions of the student. Thus, if the student is unhappy for some reason, the MR recognizes it and in collaboration with TC, it defines the suitable pedagogical actions that decrease this negative mood and try to change it into a positive one. Finally, the ER subcomponent is in every moment aware of the student's emotions during the learning process, following the forthcoming method.

So as to deal effectively with the emotions elicitation process, the Emotional Component has an affective student model where the affective information is stored [12]. An ontology of emotions is used for the formal representation of emotions. Ontology is a technique of describing formally and explicitly the vocabulary of a

domain in terms of concepts, classes, instances, relations, axioms, constraints and inference rules. It is a formal way to represent the specific knowledge of a domain, providing an explicit and extensive framework to describe it [13]. Our ontology has been built to recognize 10 emotions which are: joy, satisfaction, pride, hope, gratification, distress, disappointment, shame, fear, reproach [14]. The former five emotions compose the classification of positive emotions and are related to the positive student's emotional state. The latter five emotions compose the classification of negative emotions and are related to the negative student's emotional state. The construction of the ontology was based on the OCC cognitive theory of emotions. Thus, the concepts of the ontology are defined in terms with this theory. For instance, the positive student's emotional state and the emotion of joy are described as follows:

(POSITIVE-EMOTIONAL-STATE
(SUBCLASSES
(VALUE (JOY, SATISFACTION, PRIDE, HOPE, GRATIFICATION)))
(IS-A
(VALUE (EMOTIONAL-EVENT)))
(DEFINITION
(VALUE ("emotions or states, regarded as positive, such as joy, satisfaction, pride, hope, gratification"))))

In this way, the formal and flexible representation of an emotion can be efficiently achieved in relation to the learning goal of a student. The proposed ontology of emotions was implemented with the Protégé tool. Furthermore, we adopt an approach based on Bayesian Networks in order to extract information from the proposed "emotional" ontology and to make inferences about the emotions of the student [14]. This approach, which is used for carrying out the representation and the inference of emotions is based on the OCC model which combines the appraisal of an Event with the Intentions and Desires of a subject. Thus, taking advantage of this model, MENTOR infers about the student's emotions after the occurrence of an educational event which is related to his learning goal.

3.2 Providing the student with the appropriate affective tactic

As it has already been stated, the objective of the MENTOR is to foster the appropriate affective conditions, since these are a crucial factor for the learning process and to obtain the student with the suitable learning method. The latter goal is achieved by the Teaching Component which is responsible for providing the student with the appropriate affective tactic considering his emotional state. It consists of two subcomponents, the Teaching Generator and the Pedagogical Generator, which are responsible respectively for the appropriate teaching and pedagogical strategy as illustrated in Figure 1.

The Teaching Generator is a sub-component which is responsible for the selection and the presentation of the suitable educational material, according to the student model. The student model provides information about the cognitive status of the student such as his learning style, the knowledge that has already been acquired and his learning preferences and goals. Evaluating this information the Teaching Generator decides about the sequence of the educational material, if a theoretical or practical subject will be presented next to the student and what kind would this be, for example a more or less detailed theoretical topic or an easier or a trickier exercise [15]

The Pedagogical Generator is a sub-component which is responsible for the formation of the pedagogical actions which will be taken into account during the learning process. Once the recognition of the student's emotions and his emotional state has been stored in the affective student model, the Pedagogical Generator has all the necessary information in order to support and motivate the student to the direction of the achievement of his learning goals. As a teacher does in the real class, the Pedagogical Generator encourages the student, gives him positive feedback, congratulates him when he achieves a goal, and keeps him always in a positive mood, with the view of engaging him effectively in the learning process [16].

Combining the interaction of its two sub-components, the Teacher Component forms the appropriate affective tactic for the student. In this way, a traditional instructional tactic is enhanced with a motivational one and this would be proved beneficial to the student from two aspects. The first concerns the planning of the teaching strategy and the educational content, which and what topic will be taught to the student next and which method will be used for it. The second is more related to the delivery planning, how this topic will be taught. The role of the Pedagogical Generator, however, is not restricted only to the reassurance of the appropriateness of the teaching method or the educational material. It is concentrated also on providing the student with encouraging actions in order to preserve his positive emotional state. The pedagogical actions which have been implemented are shown in Table 1.

Table 1. The pedagogical actions of the Pedagogical Generator

Ask for giving some help	Explain the need for help
Give Help to student	Reassure the appropriateness of help
Express satisfaction after a successful help	Express unhappiness after an unsuccessful help and ask for trying again
Give explanations in an appropriate way	Express sympathy in case of fail
Encourage the student	Congratulate the student
Praise the student	Express admiration for the student
Reinforce student's efforts	Play a game with student
Give hope	Open a dialogue with the student
Play a music video clip	Present a part of a movie
Present a photo	Tell a joke

The main concern of the Teacher Component, as it is already mentioned, is to ensure that the student's mood is positive every time. To achieve this, the Teacher Component has to be aware of the student's emotions. The input that comes from the Emotional Component, which is in charge of the detection of the student's motivational state, is evaluated appropriately and thereafter the Teacher Component adapts his reaction adequately to motivate the student either by encouraging him or by praising him and in every case sustain his disposition flourishing [16]. Once the Teacher Component is aware of the student's emotions, it can proceed into the selection of the proper affective tactic.

4 Evaluation

The evaluation of the MENTOR is deployed in two axes. The first concerns the impact of the system in the learning process of students. The second examines the accuracy of MENTOR's Affective Module prediction and is concentrated on the suitability of the suggested affective tactics. Thus, we have the ability to identify factors that make the affective dimension of the educational material beneficial for

learning. We hypothesized that the enhancement of the traditional educational material with the affective dimension of MENTOR's learning framework leads to higher learning performance. In addition, we assumed that the more accurate is the prediction of the student's affective state the better is the appropriateness of the selected affected tactic and the greater is the advantage of learning with the MENTOR's educational environment in comparison to the environment of the traditional adaptive educational system.

4.1 The Framework of the Evaluation Study

4.1.1 Participants of the Experiments

In order for the process of sample gathering to be the appropriate the following procedure has been followed. A sample of 120 students in the field of computer science has been selected. Their age was between eighteen and twenty-five years old (M=20.9, SD=2.27). The students were assigned with a questionnaire containing items relative to the field of Artificial Intelligence. From the statistical evaluation of their answers a group of 108 out of the initial 120 was selected. The criterion was the lowest average which signifies the lowest starting knowledge on the domain and the lowest possible dispersion around it (M=6.24, SD=3.18). In this way, a homogeneous group was formed with the same average a-priori knowledge about the learning domain. In total, there were 65 male and 43 female students. They were randomly divided into two groups, the experimental group A and the control group B with 54 students in each group. The students of group A make use of the affective version of MENTOR, while the students of group B with the normal version of the adaptive system.

4.1.2 Questionnaires

To evaluate the student's acquired knowledge after the interaction with the system as well as the appropriate and accurate operation of the MENTOR's Affective Module, the following three questionnaires were built and proposed:

Pre-test questionnaire:

The students of the experimental and the control group attended an individual test. The test was formed by 20 items in order to check the starting knowledge of the sample. The main aim of the test was to measure the initial knowledge of each participant in the field of Artificial Intelligence. It was designed as a set of multiple choice and true/false items. A domain expert and an instructional designer contributed to the structure of the test. An item example was: "Alan Turing proposed a test that has served as a benchmark in measuring progress in the field of artificial intelligence?" In the pre-test, the students additionally asked for biographic data (for example age or sex) and for a subjective rating regarding the participants' expertise in the domain of Artificial Intelligence. The answers of each student are analyzed by statistics methods and the results compared with the post-test.

Post-test questionnaire:

The students of the experimental group A and the control group B employed different methods of learning during their interaction with the MENTOR. The students of group A made use of the affective version of the web-based adaptive educational system while the students of group B did not. That is to say, group's A learning process was enriched with the affective dimension of the e-learning system but group's B did not. After a short break of 15 minutes all participants were assigned

with a post-test questionnaire, with the aim of assessing their learning performance. The post-test questionnaire, consisting of 30 items, aimed to measure the acquired knowledge after having interacted with the MENTOR, in order to verify if and how much the system itself was useful in helping learners to reach the didactic goals. The structure of the post-test questionnaire was designed, similarly to the pre-test questionnaire, as a set of multiple choice and true/false items, with the support of a domain expert and an instructional designer.

Questionnaire on the evaluation of the Affective Module Accuracy

This questionnaire which was assigned only to the students of group A, who interacted with the affective version of MENTOR, was designed to measure the students' ratings on the basic operation of the Affective Module. The objective of this questionnaire was to assess the accuracy of the MENTOR's affective state prediction and the suitability of the suggested affective tactics as well. It was formed by eight 5-point Likert Scale items and two Fill-in items.

4.1.3 The Knowledge Domain

The learning environment of the MENTOR's web-based adaptive educational system was used in order to teach beginner's topics in Artificial Intelligence. The Knowledge Domain is constituted by 15 learning nodes, each one having theory, examples, exercises and a final evaluation test. Every student had the opportunity to interact with a pre-selected course of MENTOR for 45 minutes.

4.2 Affective versus non-affective version comparison

The affective version of MENTOR incorporates the Affective Module, while the non-affective version deprives of the Affective Module, that is operates only as a traditional web-based adaptive educational system. Taking this sceptical into account, the whole evaluation experiment was designed focused on whether the MENTOR's Affective Module and its affective learning model can improve the effectiveness of the students' study. There were two groups A and B in experiments. Group A was used as the experimental group and group B was used as the control group. The experiment started with the pre-test questionnaire. The students of two groups were asked to complete the pre-test using pencil and paper. The achieved score of each student was recorded. The results were subjected to a statistical analysis which is shown in Table 2. The pre-test results demonstrate that both groups had similar knowledge level on the knowledge domain of Artificial Intelligence.

After completing the pre-test, students interacted with the system. The group A deal with the educational material of the affective version of MENTOR, while the group B with the educational material of the non-affective version of the web-based adaptive educational system. At this phase the students had the chance to practice and enhance their knowledge dealing with the educational material via the system's interface. During this interaction, the students' actions were recorded in the system's log files, with the aim of providing information concerning the time they spent and their performance. Afterwards, the student's were assigned with a post-test and were asked to complete it using again pencil and paper. Also, the results of the post-test were subjected to a statistical analysis that is shown in Table 2.

4.2.1 The Research Question

The value of the proposed affective-learning model can be evaluated by the differences between the two groups' real study effectiveness. We consider the study effectiveness primary from the aspect of the test score, so the research question (RQ) is formulated as:

RQ: Do students interacting with the affective version of MENTOR achieve better learning results than students interacting without the affective version of MENTOR?

4.2.2 Group Analysis

In order to answer the research question we performed the analysis of the statistical differences between groups by means of the two-sample independent *t*-test. Based on the research question the null and the alternative hypotheses are formed as follows:

Null Hypothesis H_0: There is no difference, between the experimental group A and the control group B, after the interaction with the learning environment of the two versions of MENTOR.

Alternative Hypothesis H_a: The two groups A and B are different in terms of the learning performance after the interaction with the learning environment of the two versions of MENTOR.

Since a preliminary Levene's test for equality of variances indicated that the variances of the two groups were not significantly different, a two-sample *t*-test was performed that does assume equal variances, defining the significance level at α=0.05. The provided results are presented in Table 2. By analyzing furthermore the results from this table we realize that the mean grades of the students who interact with the affective version of MENTOR (EG) (M = 74.67, SD = 6.70, N = 54) was significantly different from these using the non-affective version of MENTOR (CG) (M = 71.39, SD = 6.90, N = 54), $t_{(106)}$ = 2.26, p = 0.026, where p <0.05 is considered to be significant. The standard deviation is small which means all the students expressed a consensus opinion. Moreover, the standard deviation is almost the same for both groups while the small standard error indicates that our sample means are similar to the population mean and therefore is likely to be an accurate reflection of the population. In this way, we can draw the conclusion that we can reject the non-difference null hypothesis H_0 between the experimental group A and the control group B and to accept the alternative hypothesis H_a. Namely, the group of students who interacted with the affective version of MENTOR demonstrated significant improvement in their learning progress comparing to the students who interacted with the non-affective version of MENTOR.

Table 2. Statistical Results of grade from the pre / post-test and t-Test Results for the Difference between Experimental and Control Group Means

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
grade	Equal variances assumed	,968	,328	2,261	106	,026	3,2795	1,4507	,3125	6,2579
	Equal variances not assumed			2,261	105,123	,026	3,2795	1,4507	,3122	6,2581

Group Statistics (Pre-test)

	group	N	Mean	Std. Deviation	Std. Error Mean
grade	EG	54	39,241	2,6913	,3662
	CG	54	42,185	4,6177	,6284

Group Statistics (Post-test)

	group	N	Mean	Std. Deviation	Std. Error Mean
grade	EG	54	74,673	6,7047	1,0108
	CG	54	71,393	6,9025	1,0406

From the above analysis, it is also apparent that both experimental and control groups have significant improvement in the post-test in relation to the pre-test, as it shown in Table 2. Based on this, it could be assumed that MENTOR supported successfully the students in their learning process improving their performance. Nevertheless, after the interaction with the system, the results of the post-test indicate clearly that the group A achieved better results than the group B. Consequently, it is evident that the improvement in learning of the group's A students is greater than group's B students. Therefore, it is reliable to presume that the interaction with MENTOR contributes to the student's study process significantly.

4.3 The evaluation of the Affective Module Accuracy

During the interaction with the MENTOR the students had the opportunity to deal with the educational events of MENTOR's learning environment and had been provided with various affective tactics. Every student's action was recorded by the system's log files. Consequently, in every moment MENTOR was aware of the emotional state of the student in order to offer him the appropriate affective tactic. In this phase of evaluation participated, only the students of the experimental group A, which had interacted with the MENTOR. Firstly, with the aim of recognizing the student's personality, they were given with the NEO-PI-R personality test [17], which was presented to them via the MENTOR. According to this test, MENTOR classified twenty-one students who belonged to the Extraversion category, fourteen to the Agreeableness category, nine to the Conscientiousness category, six to the Openness category and four to the Neuroticism category. In this way, students formed five groups and they were asked to fill in the questionnaire on the evaluation of the Affective Module Accuracy.

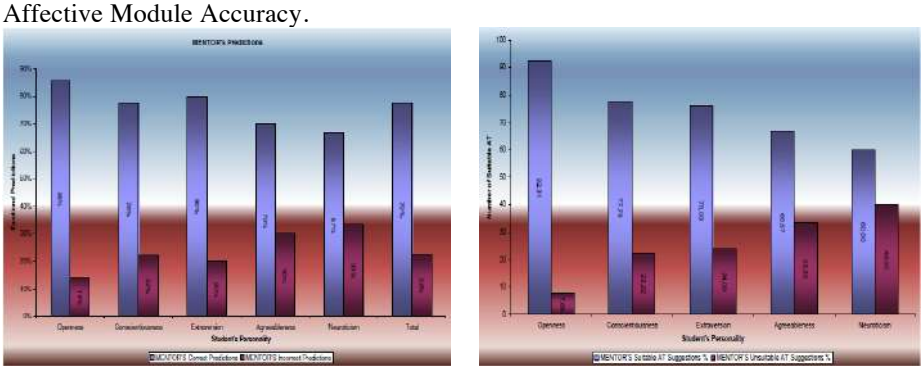


Fig.6. (a) MENTOR's prediction accuracy of students' emotions according to their personality, (b) The accuracy of MENTOR's Affective Tactic suggestions according to the students' personality.

The objective of the experiment was to measure the accuracy of the MENTOR's affective state prediction and the suitability of the suggested affective tactics as well. In order to achieve this, the students asked to declare their affective state before and after they obtained the affective tactics. Also, they asked to state if the suggested affective tactics helped them to preserve a positive mood and to achieve their educational goal. Finally, the students were given the evaluation questionnaire to fill in where they wrote down their impressions from the interaction with MENTOR. The questionnaire examined two factors, which were the students' opinion in relation to

the prediction accuracy of MENTOR as well as the appropriateness of the provided affective tactics in relation to the impact in their learning process. Then taking the students' responses into consideration and after the examination of the system's log files, we were provided with the results that are shown in the Figure 6. In this figure the categories of the students' personalities and the related system's prediction values as well as the corresponding suitability of the suggested affective tactics are demonstrated in a graphical way.

More specifically, from Figure 6 we can infer that the percentage of MENTOR's correct predictions is about 78%. We can also easily draw the conclusion that for the categories of Openness and Neuroticism, MENTOR had better and worse accuracy respectively in the prediction of their emotional states. In addition, it provides us with the inference that for the categories of Openness, Conscientiousness and Extraversion, MENTOR had better accuracy in the suggestion of the proper affective tactics.

5 Conclusions

The research presented the MENTOR Affective Learning Model which is responsible for inferring students' emotions and providing them with the appropriate affective tactic in distance learning, as well as its application in the field of the didactics of informatics. The main purpose of the MENTOR, except from the recognition of emotions, is to create and / or preserve a positive mood in the student, since this is a crucial factor for the learning process. Moreover, it aims at providing the system with suitable information about the personality and emotions of the student and also with appropriate pedagogical actions enhancing the student's motivation to "conquer" the intended knowledge. In MENTOR, the elicitation of emotions is based on a formal representation of emotions using an appropriately designed Ontology, the Affective Ontology which is implemented with the Protégé tool and is achieved by a BN-based method.

An experiment has been also conducted with the aim of evaluating MENTOR's performance and impact in learning process. The experimental results are encouraging for the educational value of the proposed model in the learning process. More specific, the research findings support the hypothesis that the affective version of a web-based adaptive educational system seems to have a significant effect on the students' attitude towards the tasks that they should perform in order to achieve their learning aim. Furthermore, all the participants reported that the system had a major contribution to their learning gain and helped them to improve their problem-solving skills. These opinions validate the accuracy of the incorporated Affective Module which is the essential element of MENTOR from the aspect of dealing with the affective information. The evaluation also verified the exactitude of MENTOR's prediction in relation to the student's affective state as well as the correctness of the suggested affective tactics. As a result, the participants reported that the guidance of the system was the appropriate by supporting them with encouraging actions in order to preserve a positive mood and to achieve efficiently their educational goal, contributing in this way to a more interesting and effective study.

References

[1] Ames, C. (1992): Classroom goals, structures and student motivation. Journal of Educational Psychology, 84(3), 261–271

[2] Conati, C., Zhou., X. (2002): Modeling students' emotions from Cognitive Appraisal in Educational Games. In: 6th International Conference on ITS, Biarritz, France

[3] Oren, T.I., Ghasem-Aghaee, N. (2003): Personality Representation Processable in Fuzzy Logic for Human Behavior Simulation. In: Proceedings of the Summer Computer Simulation Conference, Montreal, PQ, Canada, July 20-24, pp. 11–18

[4] Leontidis, M., Halatsis, C. (2007). An Affective Way to Enrich Learning. In *Proceedings of the IADIS International Conference on e-Learning*, 6-8 July, Lisbon, Portugal, 32-36.

[5] Picard, R.W. (1997): Affective Computing. MIT Press, Cambridge

[6] Scherer, K. (2000): "Psychological models of emotion". In: Borod, J. (Ed.). *The neuropsychology of emotion*, Oxford/New York: Oxford University Press, pp.137-162.

[7] Costa, P.T., McCrae, R.R. (1992): Four ways five factors are basic. Personality and Individual Differences 1 13, 653–665

[8] Ortony, A., Clore, G.L., Collins, A. (1988): The Cognitive Structure of Emotions. Cambridge University Press, Cambridge

[9] Leontidis, M., Halatsis, C., Grigoriadou, M. (2008). e-learning Issues under an Affective Perspective. In F. Li et al. (Eds.): ICWL 2008, *Lecture Notes in Computer Science,* Vol. 5145, Berlin: Springer-Verlag, 27–38.

[10] Marsella, S., & Gratch, J. (2006). EMA: A computational model of appraisal dynamics. In J. Gratch, S. Marsella, & P. Petta (Eds.), *Agent construction and emotions,* (pp. 601-606). Austrian Society for Cybernetic Studies, Vienna

[11] Leontidis, M., Halatsis, C. (2009). Affective Issues in Adaptive Educational Environments. A chapter in: *Cognitive and Emotional Processes in Web-Based Education: Integrating Human Factors and Personalization*. Mourlas, C., Tsianos, N., Germanakos, P. (Eds.), IGI Global, Hershey, USA, 111-133.

[12] Leontidis, M., Halatsis, C. (2009). Integrating Learning Styles and Personality Traits into an Affective Model to Support Learner's Learning. In Spaniol, M., Qing Li, Klamma, R., and Lau, W.H. R. (Eds.): Advances in Web Based Learning. *Lecture Notes in Computer Science*, Vol. 5686. Berlin: Springer-Verlag 225-234.

[13] Aroyo, L, Dicheva, D., Cristea, A. (2002) Ontological Support for Web Courseware Authoring, Int. Conf. on Intelligent Tutoring Systems (ITS'02), France, 270-280.

[14] Leontidis, M., Halatsis, C., Grigoriadou, M. (2009). An Ontological Approach to Infer Student's Emotions. In Fu Lee Wang, Joseph Fong, Liming Zhang and Victor S.K. Lee (Eds.): Hybrid Learning and Education. *Lecture Notes in Computer Science*, Vol. 5685. Berlin: Springer-Verlag 89-100.

[15] Leontidis,M., Halatsis,C., Grigoriadou, M.(2011). Using an affective multimedia learning framework for distance learning to motivate the learner effectively, *International Journal of Learning Technology* Vol. 6, No.3 pp. 223 - 250.

[16] Leontidis, M., Halatsis, C. (2009). Supporting Learner's Needs with an Ontology-Based Bayesian Network. In *Proceedings of the 9th IEEE International Conf. on Advanced Learning Technologies (ICALT 2009)*, Riga, Latvia, July 15-17, 455-459.

[17] Goldberg, L. R.. International Personality Item Pool (1999): A Scientific Collaboratory for the Development of Advanced Measures of Personality and Other Individual differences, Available: http://ipip.ori.org/ipip/

Iterative methods for the numerical solution of linear systems

Maria Louka *

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
mlouka@di.uoa.gr

Abstract. The objective of this dissertation is the design and analysis of iterative methods for the numerical solution of large, sparse linear systems. This type of systems emerges from the discretization of Partial Differential Equations. Two special types of linear systems are studied. The first type deals with systems whose coefficient matrix is two cyclic whereas the second type studies the augmented linear systems. Initially, the Preconditioned Simultaneous Displacement (PSD) method, which is a generalized version of the Symmetric SOR (SSOR) method, is studied when the Jacobi iteration matrix is weakly cyclic and its eigenvalues are all real "real case" or all imaginary "imaginary case". The first result is that the PSD method has better convergence rate than the SSOR method. In particular, in the "imaginary case" its convergence is increased by an order of magnitude compared to the SSOR method. In an attempt to further increase the convergence rate of the PSD method, more parameters were introduced. The new method is called the Modified PSD (MPSD) method. Under the same assumptions the convergence of the MPSD method is studied. It is shown that the optimum MPSD method is equivalent to the optimum MSOR method. Furthermore, the convergence analysis of the Generalized Modified Extrapolated SOR (GMESOR) and Generalized Modified Preconditioned Simultaneous Displacement (GMPSD) methods is studied for the numerical solution of the augmented linear systems. The main result of our analysis is that both methods possess the same rate of convergence and less complexity than the Preconditioned Conjugate Gradient (PCG) method. The last result is important since it proves that the addition of parameters in an iterative method has the same effect in the increase of the rate of convergence as that of the Conjugate Gradient (CG) method which belongs to the Krylov subspace methods.

1 Introduction

The modeling of many scientific problems leads to the solution of Partial Differential Equations (PDEs). The discretization of a PDE using finite difference or finite element methods leads to a linear system of equations whose coefficient

* Dissertation Advisor: Nikolaos Missirlis, Professor

matrix is large and sparse. These systems can be solved using direct or iterative methods. However, iterative methods become more attractive since they are very effective and require less memory and arithmetic operations than direct methods. Another reason that the iterative methods have become particularly popular is because they are suitable for parallel processing.
The first iterative methods were the Jacobi (1824) and later the Gauss-Seidel (1848). After about 100 years the popular Successive Overrelaxation (SOR) method was discovered and 10 years later, its symmetric version, the Symmetric Successive Overrelaxation (SSOR) method. These methods introduce a parameter ω whose role is to minimize the spectral radius, the largest in modulus eigenvalue, of their iterative matrix. The main result from the convergence analysis of the SOR method was the determination of the optimal value of the parameter ω for which the spectral radius is minimal and hence the rate of convergence of the iterative method becomes maximum and better, by an order of magnitude, than the Gauss-Seidel (GS) method. This result was found by Young (1952) [12] and first presented in his thesis. The whole theory was developed for systems whose coefficient matrix is two-cyclic. It was already known in 1952 that the introduction of parameters in iterative methods resulted in increasing the rate of convergence. However, research was directed to the development of other iterative methods based on orthogonality of vectors for solving generalized linear systems. A representative of these methods is the Conjugate Gradient (CG) method [6].
In this dissertation the Preconditioned Simultaneous Displacement (PSD) method is studied [8]. This method was proposed in 1980 by Evans and Missirlis [4] and is a generalization of the SSOR method for the numerical solution of linear systems. Our starting point is the derivation of a functional equation which relates the eigenvalues of the PSD preconditioned matrix to its associated Jacobi iteration matrix. In particular, convergence conditions and optimum values of the parameters of the PSD method are determined to achieve optimal rate of convergence in cases where the Jacobi iteration matrix is weakly cyclic and its eigenvalues are either all real "real case" or all imaginary "imaginary case".
The study of convergence of the PSD method revealed that its rate of convergence is faster than the SSOR method. Especially, in the "imaginary case" its convergence is improved by an order of magnitude as compared to SSOR. This result is quite encouraging for the study of the PSD method in case where the Jacobi iteration matrix has complex eigenvalues. In an effort to further increase the rate of convergence of the PSD method, more parameters were introduced. The new method called Modified PSD (MPSD)[9]. Under the same assumptions the convergence of the MPSD method is studied. The main result of this analysis is that the MPSD method becomes equivalent to the MSOR method for the optimum values of their parameters. It is also shown that the MPSD method converges faster than the corresponding Modified SSOR method. Also, the PSD method achieves faster convergence rate compared to the classical SOR method. Indeed, in case where the smallest in modulus eigenvalue of the Jacobi iteration matrix increases then the rate of convergence of the MPSD method increases

whereas the rate of convergence of the SOR method remains constant.

In recent years, many researchers have studied the saddle point problem which leads to the solution of an augmented linear system. Such systems arise in areas of computational fluid dynamics, constrained optimization, image processing, finite element approximations and elsewhere. The most known and the oldest methods are the Uzawa and the preconditioned Uzawa methods which are special cases of the SOR-like method. In 2005, the Generalized SOR (GSOR) method was studied, which improved the rate of convergence of the SOR-like method by introducing an additional parameter.
In this dissertation, we developed the convergence analysis of the Generalized Modified Extrapolated SOR (GMESOR) and generalized Modified Preconditioned Simultaneous Displacement (GMPSD) methods [10] for the numerical solution of augmented linear systems. To study the convergence of these methods it was necessary to derive a functional equation between the eigenvalues of the iteration matrices of the aforementioned methods with those of matrix J (see (11)). It is assumed that the eigenvalues of the matrix J are all real and positive. Under these assumptions, sufficient conditions for the convergence of these methods are found. Furthermore, the optimal values of their parameters are determined such that these methods obtain the optimum rate of convergence. We studied the Generalized SOR (GSOR), Generalized Extrapolated SOR (GESOR), Generalized Modified PSD with three parameters (GMPSD(3)), Generalized Modified SSOR (GMSSOR) and Generalized SSOR (GSSOR) methods. The main result of this analysis is that all these methods have the same rate of convergence and less complexity than the Preconditioned Conjugate Gradient (PCG) method. The latter result is important because it demonstrates that the introduction of parameters in an iterative method results in the same increase in the rate of convergence as the Conjugate Gradient (CG) method. Next, we present a small part of the present dissertation, which refers to the convergence analysis of the Generalized Modified Extrapolated SOR method.

2 The Generalized Modified Extrapolated SOR method

Let $A \in \mathbb{R}^{m\times m}$ be a symmetric positive definite matrix and $B \in \mathbb{R}^{m\times n}$ be a matrix of full column rank, where $m \geq n$. Then, the augmented linear system is of the form [1], [2], [3], [5]

$$\mathcal{A}u = b \tag{1}$$

where

$$\mathcal{A} = \begin{pmatrix} A & B \\ -B^T & 0 \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ -b_2 \end{pmatrix} \tag{2}$$

with B^T denoting the transpose of the matrix B. Such systems arise in areas of computational fluid dynamics, constrained optimization, image processing , in finite element approximations and elsewhere [2].
Let the coefficient matrix $\mathcal{A}$ of (1) be defined by the splitting

$$\mathcal{A} = \mathcal{D} - \mathcal{L} - \mathcal{U} \tag{3}$$

where

$$\mathcal{D} = \begin{pmatrix} A & 0 \\ 0 & Q \end{pmatrix},\ \mathcal{L} = \begin{pmatrix} 0 & 0 \\ B^T & aQ \end{pmatrix},\ \mathcal{U} = \begin{pmatrix} 0 & -B \\ 0 & (1-a)Q \end{pmatrix}, \tag{4}$$

with $Q \in \mathbb{R}^{n\times n}$ be a prescribed nonsingular and symmetric matrix and $a \in \mathbb{R}$. Furthermore, we denote by T, the diagonal matrix $T = diag(\tau_1 I_m, \tau_2 I_n)$ with $\tau_1, \tau_2 \in \mathbb{R} - \{0\}$, $I_m \in \mathbb{R}^{m\times m}$ and $I_n \in \mathbb{R}^{n\times n}$ be identity matrices.
For the numerical solution of (1), we consider the following iterative scheme

$$\begin{pmatrix} x^{(k+1)} \\ y^{(k+1)} \end{pmatrix} = \mathcal{H}(\tau_1, \tau_2) \begin{pmatrix} x^{(k)} \\ y^{(k)} \end{pmatrix} + \eta(\tau_1, \tau_2) \begin{pmatrix} b_1 \\ -b_2 \end{pmatrix} \tag{5}$$

where

$$\mathcal{H}(\tau_1, \tau_2) = I - R^{-1}T\mathcal{A},\ \ \eta(\tau_1, \tau_2) = R^{-1}Tb, \tag{6}$$

R is a nonsingular matrix to be defined and $I = diag(I_m, I_n)$.
In the sequel we consider different types of the preconditioned matrix R and study the iterative methods derived by (5) and (6).

2.1 The functional relationship

As a first step we consider the preconditioning matrix which is formed by the parametrized diagonal and lower triangular part of $\mathcal{A}$

$$R = \mathcal{D} - \Omega\mathcal{L}, \tag{7}$$

where $\Omega = diag(\omega_1 I_m, \omega_2 I_n)$ with $\omega_1, \omega_2 \in \mathbb{R}$. Then the iterative scheme (5), (6) becomes the GMESOR method. In case $a = 0$ this method was introduced in [1] and proposed for further study. We initiate our study by developing the convergence analysis of GMESOR. In the general case where $a \neq 0$ our theoretical analysis reveals new convergence regions for the parameters of the GSOR method generalizing the ones found in [1]. If R is given by (7), then (6) becomes

$$\mathcal{H}(\tau_1, \tau_2, \omega_2, a) = I - (\mathcal{D} - \Omega\mathcal{L})^{-1}T\mathcal{A}$$

or

$$\mathcal{H}(\tau_1, \tau_2, \omega_2, a) = (\mathcal{D} - \Omega\mathcal{L})^{-1}[(I - T)\mathcal{D} + (T - \Omega)\mathcal{L} + T\mathcal{U}] \tag{8}$$

and

$$\eta(\tau_1, \tau_2, \omega_2, a) = (\mathcal{D} - \Omega\mathcal{L})^{-1}Tb. \tag{9}$$

The iterative scheme given by (5), (8) and (9) will be referred to as the Generalized Modified Extrapolated SOR (GMESOR) method. For $(\mathcal{D} - \Omega\mathcal{L})^{-1}$ to exist we require

$$\det(\mathcal{D} - \Omega\mathcal{L}) \neq 0.$$

Because of (4)

$$R = \mathcal{D} - \Omega\mathcal{L} = \begin{pmatrix} A & 0 \\ -\omega_2 B^T & (1 - a\omega_2)Q \end{pmatrix}.$$

Therefore,

$$\det(\mathcal{D} - \Omega\mathcal{L}) = (1 - a\omega_2)^n \det A \det Q \neq 0$$

or

$$a\omega_2 \neq 1 \tag{10}$$

since the matrix A is symmetric positive definite and the matrix Q is nonsingular.
The GMESOR method has the following algorithmic form.

THE GMESOR METHOD: *Let $Q \in \mathbb{R}^{n\times n}$ be a nonsingular and symmetric matrix. Given initial vectors $x^{(0)} \in \mathbb{R}^m$ and $y^{(0)} \in \mathbb{R}^n$, and the parameters $\tau_1, \tau_2 \neq 0,\ \omega_2, a \in \mathbb{R}$ with $a\omega_2 \neq 1$. For $k = 0, 1, 2, ...$ until the iteration sequence $\{({x^{(k)}}^T, {y^{(k)}}^T)^T\}$ is convergent, compute*

$x^{(k+1)} = (1 - \tau_1)x^{(k)} + \tau_1 A^{-1}(b_1 - By^{(k)})$,
$y^{(k+1)} = y^{(k)} + \frac{1}{1-a\omega_2} Q^{-1}\left\{B^T[\omega_2 x^{(k+1)} + (\tau_2 - \omega_2)x^{(k)}] - \tau_2 b_2\right\}$,

where Q is an approximate (preconditioning) matrix of the Schur complement matrix $B^T A^{-1} B$.

Note that in the above algorithm the parameter ω_1 is eliminated. For special values of its parameters GMESOR degenerates into known methods or produces new ones. Indeed, if $\omega = \tau_1 = \tau_2 = \omega_2$ and $a = 0$ then GMESOR becomes the SOR-like method [5]; if $\omega = \tau_1 = \tau_2 = \omega_2 = 1$ and $a = 0$ then it becomes the preconditioned Uzawa method [3]; if $\tau = \tau_1 = \tau_2$ then GMESOR will be referred to as the GESOR method and if $\tau_1 = \omega_1,\ \tau_2 = \omega_2$ and $a = 0$, then it becomes the GSOR method [1].
By comparing the algorithmic structures of the GMESOR method and the GSOR method, we can verify that both methods have exactly the same computational complexity. Also, the GMESOR method has less computational complexity than the Preconditioned Conjugate Gradient (PCG) method.
In the following theorem we find the functional relationship between the eigenvalues λ of the iteration matrix $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ with the eigenvalues μ of the associated matrix J, where

$$J = Q^{-1}B^T A^{-1}B. \tag{11}$$

Theorem 1 *Let $A \in \mathbb{R}^{m\times m}$ be symmetric positive definite, $B \in \mathbb{R}^{m\times n}$ be of full column rank and $Q \in \mathbb{R}^{n\times n}$ be nonsingular and symmetric. If $\lambda \neq 1 - \tau_1$ is an eigenvalue of the matrix $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ and if μ satisfies*

$$\lambda^2 + \lambda\left(\tau_1 - 2 + \frac{\tau_1\omega_2}{1 - a\omega_2}\mu\right) + 1 - \tau_1 + \frac{\tau_1(\tau_2 - \omega_2)}{1 - a\omega_2}\mu = 0, \tag{12}$$

where $a\omega_2 \neq 1$, then μ is an eigenvalue of the key matrix $J = Q^{-1}B^T A^{-1}B$. Conversely, if μ is an eigenvalue of J and if $\lambda \neq 1 - \tau_1$ satisfies (12), then λ

is an eigenvalue of $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$. *In addition,* $\lambda = 1 - \tau_1$ *is an eigenvalue of* $\mathcal{H}(\tau_1, \tau_2, \omega_2, a)$ *(if* $m > n$*) with the corresponding eigenvector* $(x^T, 0)^T$, *where* $x \in \mathcal{N}(B^T)$ *and* $\mathcal{N}(B^T)$ *is the nullspace of* B^T.

Proof. See [10].
From the above theorem we can obtain the following corollary.

Corollary 1 *Under the hypothesis of Theorem 1 the nonzero eigenvalues of the iteration matrix* $\mathcal{L}(\omega_1, \omega_2, a)$ *of the GSOR method are given by* $\lambda = 1 - \omega_1$ *or if* $a\omega_2 \neq 1$ *by*

$$\lambda^2 + \lambda \left(\omega_1 - 2 + \frac{\omega_1 \omega_2}{1 - a\omega_2} \mu \right) + 1 - \omega_1 = 0. \tag{13}$$

2.2 Optimum parameters

In this section we determine optimum values for the parameters of the GSOR and GMESOR methods under the hypothesis that $a \neq 0$ and the eigenvalues of the matrix J are real. The sign of J's eigenvalues depends upon the properties of the matrix Q. We assume that Q is a symmetric positive definite matrix. The matrix Q is an approximate matrix to $B^T A^{-1} B$. The reason being that if $Q \simeq B^T A^{-1} B$ then $J = Q^{-1} B^T A^{-1} B \simeq I$. In this case the ratio of the maximum to the minimum eigenvalue of the matrix J becomes minimum and its value is approximately 1. As a consequence, the spectral radius of the iteration matrix of the GMESOR (GSOR) method attains its minimum value.

The GSOR method

In the following theorem the optimum parameters for the GSOR method are determined assuming that $a \neq 0$.

Theorem 2 *Consider the GSOR method. Let* $A \in \mathbb{R}^{m \times m}$ *and* $Q \in \mathbb{R}^{n \times n}$ *be symmetric positive definite and* $B \in \mathbb{R}^{m \times n}$ *be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix* $J = Q^{-1} B^T A^{-1} B$ *by* μ_{min} *and* μ_{max}, *respectively. Then the spectral radius of the GSOR method,* $\rho(\mathcal{L}(\omega_1, \omega_2, a))$, *is minimized for any* $a \neq -\sqrt{\mu_{min}\mu_{max}}$ *at*

$$\omega_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad \text{and} \quad \omega_{2_{opt}} = \frac{1}{a + \sqrt{\mu_{min}\mu_{max}}} \tag{14}$$

and its corresponding value is

$$\rho(\mathcal{L}(\omega_{1_{opt}}, \omega_{2_{opt}}, a)) = (1 - \omega_{1_{opt}})^{\frac{1}{2}} = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{15}$$

Proof. The functional relationship (13) may be written as follows

$$(\lambda + \omega_1 - 1)(\lambda - 1) = -\lambda \omega_1 \hat{\omega}_2 \mu \tag{16}$$

where

$$\hat{\omega}_2 = \frac{\omega_2}{1 - a\omega_2}, \tag{17}$$

and $a\omega_2 \neq 1$. The optimum values of ω_1 and $\hat{\omega}_2$ will be determined such that

$$\rho(\mathcal{L}(\omega_1, \hat{\omega}_2, a)) = \max_{\mu_{min} \leq \mu \leq \mu_{max}} |\lambda| \tag{18}$$

is minimum. Then, the real roots of (16) are the intersection points of the parabola

$$g_{\omega_1}(\lambda) = \frac{(\lambda + \omega_1 - 1)(\lambda - 1)}{\omega_1 \hat{\omega}_2} \tag{19}$$

and the straight lines

$$h(\lambda) = -\lambda\mu, \;\; 0 < \mu_{min} \leq \mu \leq \mu_{max}. \tag{20}$$

Following a similar argument as in [11] page 111, $h(\lambda)$ are straight lines through the point $(0, 0)$ and $g_{\omega_1}(\lambda)$ is a parabola passing through the point (1,0). The discriminant of (13) is

$$\Delta(\omega_1, \; \hat{\omega}_2, \; \mu) = (2 - \omega_1 - \omega_1\hat{\omega}_2\mu)^2 - 4(1 - \omega_1) = 0. \tag{21}$$

Note that $\Delta(\omega_1, \hat{\omega}_2, \mu) \leq 0$ for $0 < \omega_1 \leq \tilde{\omega}$ and $\Delta(\omega_1, \hat{\omega}_2, \mu) \geq 0$ for $\tilde{\omega} \leq \omega_1 < 2$, where

$$\tilde{\omega} = \frac{4\hat{\omega}_2\mu}{(1 + \hat{\omega}_2\mu)^2}.$$

If $0 < \omega_1 \leq \tilde{\omega}$ then the minimum value of $\rho(\mathcal{L}(\omega_1, \hat{\omega}_2, a))$ is attained when (see (13))

$$|\tilde{\lambda}_1| = |\tilde{\lambda}_N| = (1 - \omega_1)^{1/2}, \tag{22}$$

where $\tilde{\lambda}_1$ and $\tilde{\lambda}_N$ are the two conjugate complex roots of (13) as illustrated in figure 1. Furthermore, (22) is a decreasing function of ω_1. In case $\tilde{\omega} \leq \omega_1 < 2$ the roots of (13) can be geometrically interpreted as the intersection of the curves $g_{\omega_1}(\lambda)$ and $h_1(\lambda) = -\lambda\mu_{max}$. The largest abscissa of the two points of intersection of $h_1(\lambda)$ and $g_{\omega_1}(\lambda)$ decreases with increasing ω_1. Indeed as ω_1 increases, the intersection point $(1-\omega_1, 0)$ of $g_{\omega_1}(\lambda)$ with the $O\lambda$ axis is moving towards to zero until $g_{\omega_1}(\lambda)$ becomes tangent to $h_1(\lambda)$, which occurs when $\Delta(\omega_1, \; \hat{\omega}_2, \; \mu_{max}) = 0$ or equivalently

$$(2 - \omega_1 - \omega_1\hat{\omega}_2\mu_{max})^2 - 4(1 - \omega_1) = 0. \tag{23}$$

A similar argument for $h_N(\lambda) = -\lambda\mu_{min}$ reveals the condition $\Delta(\omega_1, \hat{\omega}_2, \mu_{min}) = 0$ must hold or equivalently

$$(2 - \omega_1 - \omega_1\hat{\omega}_2\mu_{min})^2 - 4(1 - \omega_1) = 0. \tag{24}$$

Note that the straight lines $h_1(\lambda) = -\lambda\mu_{max}$ and $h_N(\lambda) = -\lambda\mu_{min}$ include all the lines $h(\lambda) = -\lambda\mu$. The spectral radius is given by

$$\rho(\mathcal{L}(\omega_1, \hat{\omega}_2, a)) = \max_{\mu_{min} \leq \mu \leq \mu_{max}} \{|\tilde{\lambda}_1|, |\tilde{\lambda}_N|\} \tag{25}$$

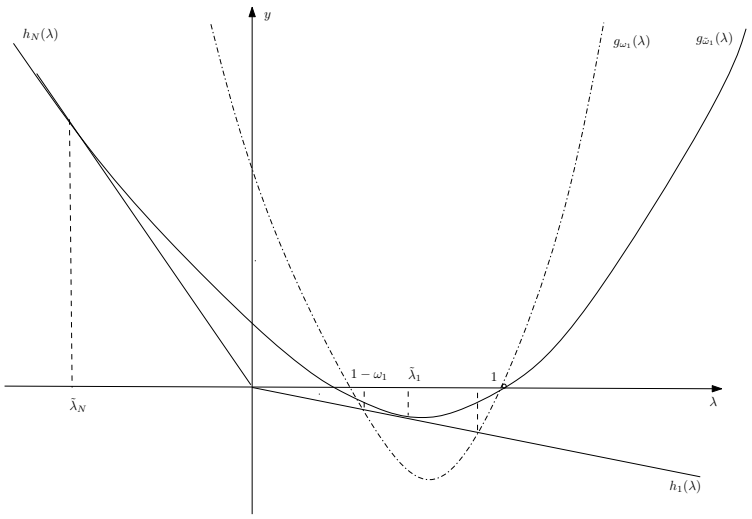


Fig. 1. Conditions for minimization of $\rho(\mathcal{L}(\omega_1, \hat{\omega}_2, a))$.

where $\tilde{\lambda}_1, \tilde{\lambda}_N$ are the abscissas of the points of tangent of $h_1(\lambda), h_N(\lambda)$, respectively. For the minimization of $\rho(\mathcal{L}(\omega_1, \omega_2, a))$ with respect to ω_1 we require

$$|\tilde{\lambda}_1| = |\tilde{\lambda}_N|$$

or

$$\tilde{\lambda}_1 = -\tilde{\lambda}_N = (1-\omega_1)^{1/2}, \tag{26}$$

where the last equality holds by the fact that $\tilde{\lambda}_1,\ \tilde{\lambda}_N$ are the abscissas of the tangents $h_1(\lambda)$ and $h_N(\lambda)$, respectively. Equating the first parts of (23) and (24) we obtain

$$2 - \omega_1 - \omega_1\hat{\omega}_2\mu_{max} = -(2-\omega_1) + \omega_1\hat{\omega}_2\mu_{min}$$

or

$$\omega_1 = \frac{4}{2 + \hat{\omega}_2(\mu_{min} + \mu_{max})}. \tag{27}$$

Substituting (27) into (23), it follows that

$$\hat{\omega}_2 = \frac{1}{\sqrt{\mu_{min}\mu_{max}}}, \tag{28}$$

from which, because of (17), the second part of (14) is obtained. From (27), because of (28), we obtain that the optimum value of ω_1, is given by the first part of (14). From (22) and (26) it follows that

$$\rho(\mathcal{L}(\omega_1, \omega_2, a)) = (1-\omega_1)^{1/2}$$

which, because of (14), yields (15). □

Theorem 2 finds the optimum values of the relaxation parameters ω_1 and ω_2 of the GSOR method in the general case where $a \neq 0$. Note that by letting $a = 0$ in (14) we obtain the optimums found in [1]. Our analysis shows that the parameter a has no impact on the spectral radius of the GSOR method as one might have expected. In fact, from (14) it follows that $\omega_{2_{opt}} \in (0, (\mu_{min}\mu_{max})^{-1/2}]$ for any $a \neq -\sqrt{\mu_{min}\mu_{max}}$. This implies that GSOR will attain the same rate of convergence for any value of ω_2 in the range $(0, (\mu_{min}\mu_{max})^{-1/2}]$. In case μ_{min} and μ_{max} cannot be estimated accurately enough this is an advantage compared to the single value $(\mu_{min}\mu_{max})^{-1/2}$ for ω_2 in the GSOR with $a = 0$.

The GMESOR method

In the sequel we determine the optimum parameters for the GMESOR method.

Theorem 3 *Consider the GMESOR method. Let $A \in \mathbb{R}^{m\times m}$ and $Q \in \mathbb{R}^{n\times n}$ be symmetric positive definite and $B \in \mathbb{R}^{m\times n}$ be of full column rank. Denote the minimum and the maximum eigenvalues of the matrix $J = Q^{-1}B^TA^{-1}B$ by μ_{min} and μ_{max}, respectively. Then the spectral radius of the GMESOR method, $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a))$, is minimized at*

$$\omega_{2_{opt}} = \tau_{2_{opt}}, \tag{29}$$

$$\tau_{1_{opt}} = \frac{4\sqrt{\mu_{min}\mu_{max}}}{(\sqrt{\mu_{min}} + \sqrt{\mu_{max}})^2} \quad \textit{and} \quad \tau_{2_{opt}} = \frac{1}{a + \sqrt{\mu_{min}\mu_{max}}} \tag{30}$$

and its corresponding value is

$$\rho(\mathcal{H}(\tau_{1_{opt}}, \tau_{2_{opt}}, \omega_{2_{opt}}, a)) = \frac{\sqrt{\mu_{max}} - \sqrt{\mu_{min}}}{\sqrt{\mu_{max}} + \sqrt{\mu_{min}}}. \tag{31}$$

Proof. The functional relationship of the GMESOR method is written as (12) or

$$(1 - a\omega_2)(\lambda + \tau_1 - 1)(\lambda - 1) = \tau_1(\omega_2 - \tau_2 - \lambda\omega_2)\mu. \tag{32}$$

The optimum values of τ_1, τ_2 and ω_2 will be determined such that

$$\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a)) = \max_{\mu_{min}\leq\mu\leq\mu_{max}} |\lambda| \tag{33}$$

is minimum. Then the real roots of (32) are the intersection points of the parabola

$$g(\lambda) = \frac{(\lambda + \tau_1 - 1)(\lambda - 1)(1 - a\omega_2)}{\tau_1} \tag{34}$$

and the straight lines

$$h(\lambda) = (\omega_2 - \tau_2 - \lambda\omega_2)\mu, \;\; 0 < \mu_{min} \leq \mu \leq \mu_{max}. \tag{35}$$

Following a similar argument as in [11] page 111, $h(\lambda)$ are straight lines through the point $(0, (\omega_2 - \tau_2)\mu)$ and $g(\lambda)$ is a parabola passing through the points (1,0)

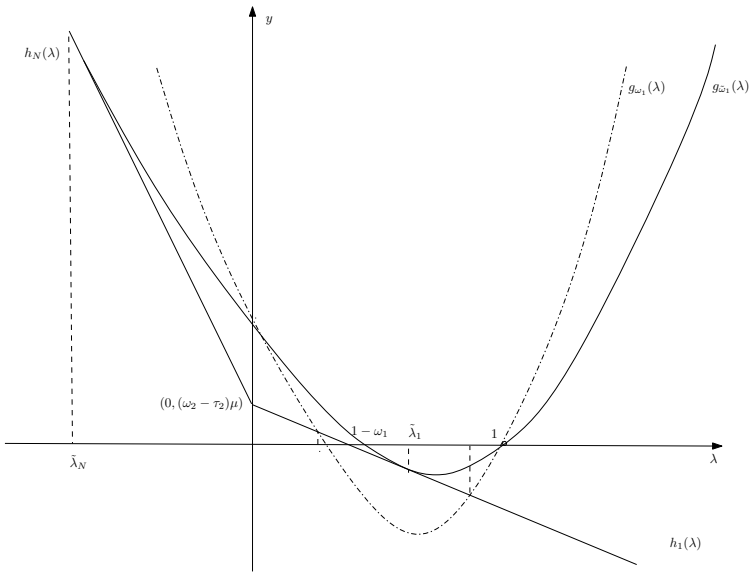


Fig. 2. Conditions for minimization of $\rho(\mathcal{H}(\tau_1,\tau_2,\omega_2))$.

and $(1-\tau_1,0)$ (see figure 2). It can be verified that the largest abscissa of the two points of intersection of $h(\lambda)$ and $g(\lambda)$ decreases until $h(\lambda)$ becomes tangent to $g(\lambda)$ which occurs when the discriminant of (12) becomes equal to zero, i.e. $\Delta(\tau_1,\tau_2,\omega_2,\mu)=0$ or

$$[(\tau_1-2)(1-a\omega_2)+\tau_1\omega_2\mu]^2-4(1-a\omega_2)[(1-\tau_1)(1-a\omega_2)+\tau_1(\tau_2-\omega_2)\mu]=0. \quad (36)$$

Note that since the straight lines $h_1(\lambda)=(\omega_2-\tau_2-\lambda\omega_2)\mu_{max}$ and $h_N(\lambda)=(\omega_2-\tau_2-\lambda\omega_2)\mu_{min}$ include all the lines $h(\lambda)=(\omega_2-\tau_2-\lambda\omega_2)\mu$, the optimum values of the parameters $\tau_1,\ \tau_2$ are obtained when $h_1(\lambda)$ and $h_N(\lambda)$ are tangent to the parabola $g_{\tau_1}(\tilde{\lambda})$. Furthermore

$$\rho(\mathcal{H}(\tau_1,\tau_2,\omega_2,a))=\max_{\mu_{min}\le\mu\le\mu_{max}}\{|\tilde{\lambda}_1|,|\tilde{\lambda}_N|\} \quad (37)$$

where $\tilde{\lambda}_1,\tilde{\lambda}_N$ are the abscissas of the points of tangent of $h_1(\lambda),h_N(\lambda)$, respectively. Therefore,

$$|\tilde{\lambda}_1|=[(1-\tau_1)(1-a\omega_2)+\tau_1(\tau_2-\omega_2)\mu_{max}]^{1/2}, \quad (38)$$

and

$$|\tilde{\lambda}_N|=[(1-\tau_1)(1-a\omega_2)+\tau_1(\tau_2-\omega_2)\mu_{min}]^{1/2}, \quad (39)$$

From (37) it follows that the minimum value of $\rho(\mathcal{H}(\tau_1,\tau_2,\omega_2,a))$ is attained when

$$|\tilde{\lambda}_1|=|\tilde{\lambda}_N| \quad (40)$$

which, because of (38) and (39), it follows that

$$\omega_2 = \tau_2. \tag{41}$$

In case $\tilde{\lambda}_1$ and $\tilde{\lambda}_N$ are the two conjugate complex roots of (32), it follows that (40) holds also. So, (41) holds if either (32) has real or conjugate complex roots. However, if (41) holds, then (12) becomes

$$\lambda^2 + \lambda\left(\tau_1 - 2 + \tau_1\hat{\tau}_2\mu\right) + 1 - \tau_1 = 0,$$

which is the functional relationship of the GSOR method (see (13)) with

$$\hat{\tau}_2 = \frac{\tau_2}{1 - a\tau_2}. \tag{42}$$

Therefore the optimum values of τ_1 and $\hat{\tau}_2$ are given by the first and second part of (14), respectively, whereas the minimum value of $\rho(\mathcal{H}(\tau_1, \tau_2, \omega_2, a))$ is given by (15). Finally, using (42) we find (30). □

In 2003, Li, Evans and Zhang [7], applied the Preconditioned Conjugate Gradient (PCG) method for solving the augmented linear system (1) and proved that the PCG method is at least as fast as the SOR-like method. Later, in [1] it was established that the GSOR method has better convergence rate than the SOR-like method whereas its spectral radius is the same with that of the PCG method for the optimum values of its parameters. Our analysis shows that the GMESOR iterative method have also the same rate of convergence with the GSOR method for the optimum values of their parameters (see Theorems 2, 3), which in turn is equal to the PCG method.

3 Remarks and Conclusions

We have studied the convergence analysis of various generalized iterative methods for the solution of the augmented linear system (1) when the coefficient matrix $\mathcal{A}$ is of the form (2). We assumed that $A \in \mathbb{R}^{m\times m}$ was a symmetric positive definite matrix and $B \in \mathbb{R}^{m\times n}$ was a matrix of full column rank, where $m \geq n$, in order to have a unique solution, whereas Q was a symmetric positive definite matrix. Under these assumptions we were able to find sufficient conditions for the GMESOR iterative method as well as for its counterparts to converge and we were able to determine its optimum rate of convergence in the general case where $a \neq 0$. From our analysis, it is proved that this method is equivalent with the GSOR method since it has the same spectral radius, which is given by (15). It is also proved that the introduction of the parameter a in the structure of $\mathcal{A}$ and hence in the preconditioned matrix R does not have any impact in the convergence rate of this method as one might have expected. Furthermore, all these methods have the same spectral radius as the Preconditioned Conjugate Gradient (PCG) method but less complexity. Therefore, it will be interesting to study the behavior of the GSOR and GMESOR methods in problems where the

PCG method is the best solver. An interesting research direction is the study of all these methods in case of nonsymmetric augmented linear systems where the eigenvalues of the matrix J are now complex.

References

1. Z.-Z. Bai, B. N. Parlett and Z.-Q. Wang, On generalized succesive overrelaxation methods for augmented linear systems, Numer. Math. 102, (2005), 1-38.
2. M. Benzi, G. H. Golub and J. Liesen, Numerical solution of saddle point problems, Acta Numerica, (2005), 1-137.
3. H. C. Elman and G. H. Golub, Inexact and preconditioned Uzawa algorithms for saddle point problems, SIAM J. Numer. Anal. 31, (1994), 1645-1661.
4. D. J. Evans and N. M. Missirlis, The preconditioned simultaneous displacement method (PSD method) for elliptic difference equations, Mathematics and Computers in Simulation 22, (1980), 256-263.
5. G. H. Golub, X. Wu and J.-Y Yuan, SOR-like methods for augmented systems, BIT 41, (2001), 71-85.
6. M. R. Hestenes and E. Stiefel, Methods of Conjugate Gradients for Solving Linear Systems, Journal of Research of the National Bureau of Standards, vol. 49, No. 6, (1952), 409-436.
7. C.-J. Li, Z. Li, D. J. Evans and T. Zhang, A note on an SOR-like method for augmented systems, IMA J. Numer. Anal. 23, (2003), 581-592.
8. M. A. Louka, N. M. Missirlis and F. I. Tzaferis, The impact of the eigenvalue locality on the convergence behavior of the PSD method for two-cyclic matrices, Lin. Alg. and its Appl., Vol. 430, No 8, pp. 1929-1944, 2009.
9. M. A. Louka, N. M. Missirlis and F. I. Tzaferis, Is modified PSD equivalent to modified SOR for two-cyclic matrices? Lin. Alg. and its Appl., Vol. 432, No. 11, pp. 2798-2815, 2010.
10. M. A. Louka and N. M. Missirlis, Preconditioning augmented linear systems (in preparation).
11. R. S. Varga, Matrix Iterative Analysis, Prentice-Hall, Inc. Englewood Cliffs, N.J., 1962.
12. D. M. Young, Iterative Solution of Large Linear Systems, Academic Press, New York, 1971.

Testing CMOS Digital ICs with Analog Techniques

Sotiris Matakias*

Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
s.matakias@di.uoa.gr

Abstract. In this thesis three novel analog techniques for testing CMOS Integrated circuits are presented. These techniques are based on analog circuits since they offer a number of important advantages compared to standard digital test techniques, such us less silicon area, lower power consumption and high operating speed. Therefore, the proposed techniques can be embedded in the circuit under test, contributing to the design of more reliable circuits.

Keywords: Self Checking Checkers, Two Rail Code Checkers, Current Mode Checker, Periodic Output Checkers. Soft Errors, Sense Amplifier, Timing Errors, I_{DDQ} testing, Current Mirror Amplifier

1 Introduction

A widely used error detection code in fault secure systems is the Two Rail Code (TRC) [3]. The first analogue technique of this thesis is a current mode, parallel TRC checker suitable for the implementation of high fan-in embedded checkers. The new circuit belongs to the periodic outputs category of TRC checkers and provides high testability since it is totally self-checking (TSC) [1] or strongly code-disjoint (SCD) [2] for a wide set of realistic faults, including transistor stuck-open faults that are not covered by other TRC checkers in the same category. Any TSC checker is capable to detect all internal faults if all codewords are available at the checkers' inputs. Designs of the proposed TRC checker, in a standard $0.18\mu m$ CMOS technology proved the efficiency of the circuit over earlier topologies in the same category, in terms of silicon area requirements, speed performance and power consumption.

A very important class of faults is the transient faults that cause soft or timing errors due to a variety of mechanisms, such as radiation, power supply noise, e.t.c. The shrinking of dimensions in CMOS technology makes digital circuits more sensitive to such mechanisms. We propose a novel and fast concurrent soft and timing error detection circuit for CMOS ICs based on current mode sense amplifier topologies. The circuit exploits the temporary nature of the transient faults as well as the delayed response of the delay faults to detect the corresponding errors.

* Dissertation Advisor: Angela Arapoyanni, Assoc. Professor.

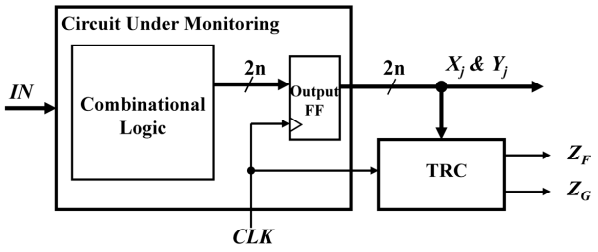


Fig. 1. A self-checking circuit with a TRC checker

Our third analogue fault detection technique is an I_{DDQ} testing technique. I_{DDQ} testing is a valuable manufacturing tool to achieve high defect detection levels and improve quality and reliability of CMOS ICs. A new I_{DDQ} testing technique, a suitable embedded circuit to support it and a theoretical model for the circuit operation are presented in this thesis. In deep submicron technologies, the discrimination between defective and non-defective I_{DDQ} currents is hard. In order to be able to exploit I_{DDQ} testing in nanometer technologies we propose a new I_{DDQ} testing approach where the background current at the sensing node is properly controlled taking into account possible process and temperature variations as well as the dependence of the background current on the applied test vector. The adoption of this method is a promising way to extend the viability of I_{DDQ} testing in the nanometer technologies.

This abstract is organized as follows. In Section 2 the proposed TRC checker is presented along with a modified version for enhanced testability. In Section 3 a new circuit for soft and timing error detection based on a sense amplifier is given and finally in Section 4 the proposed technique for I_{DDQ} testing is presented.

2 A Current Mode, Parallel Two-Rail Code Checker

A new parallel, fast and low silicon area cost TRC checker is proposed in this thesis. The new checker has periodic outputs (in each clock semi-period they have alternating complementary values) and it is based on the current mode structure we introduced in [6]. It is suitable for the implementation of embedded, high fan-in TRC checkers. The new checker is proved to be TSC or SCD for a wide set of realistic faults, while a modified version of it covers transistor stuck-open faults that are not fully detectable in earlier TRC checker designs [5]. Note that stuck-open faults present a considerable interest in very deep submicron technologies [7, 9]. In addition, like in [5], the checker requires only two input codewords, out of a wide variety of equivalent pairs, to satisfy the TSC or SCD property for the enhanced set of faults.

The general topology of a circuit that is monitored by a two-rail code (TRC) checker is shown in Fig. 1. The circuit under monitoring is designed to produce two-railed output words $(X_j, Y_j,\ j \in [1, \ldots, n])$ when it is fault-free $(X_j = \overline{Y}_j)$ and non two-railed output words $(X_j = Y_j)$ in case of internal faults.

The proposed n-variable TRC checker is presented in Fig. 2. The circuit is divided into two identical sub-blocks, the F-SubBlock (FSB) and the G-SubBlock (GSB); it receives n pairs of two-railed inputs $(X_j, Y_j, j \in [1, \ldots, n])$ and provides a two-railed pair of outputs Z_F and Z_G, one for each sub-block. Since this checker belongs to the periodic outputs TRC checkers category, it has been designed so that the outputs Z_F and Z_G present alternating complementary logic values in each semi-period of the clock signal. The first sub-block is fed by half of the checker input pairs $(X_r, Y_r, r \in [1, \ldots, k]$, where $k = n/2)$ and the complementary clock signal $CLKB$ while the second is fed by the rest of the input pairs $(X_s, Y_s, s \in [k+1, k+2, \ldots, n])$ and the clock signal CLK.

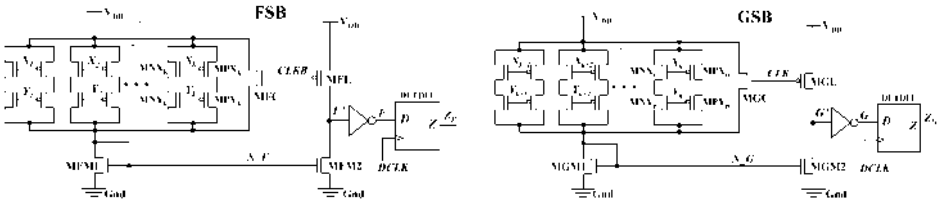


Fig. 2. The proposed two-rail code checker, $k = n/2$

The checker's outputs Z_F and Z_G always present complementary logic values in the fault free operation of the circuit under monitoring and non-complementary in the opposite case. The checker operation is described in [10] and is transparent to the circuit under monitoring.

The waveforms in Fig. 3 show the response of the checker's nodes F and G in the presence of codeword inputs and all possible non-codeword input conditions. In all three cases, the Z_F and Z_G outputs of the checker will capture the responses on F and G indicating the presence of errors or not and the proposed circuit is proved to be code-disjoint.

It is proved [11] that the proposed checker is TSC for the following kind of faults: line stuck-at faults, Transistor Stuck-On (TSON) faults, transient faults, Transistor Stuck Open (TSOP) faults (except for the 4 input transistors) and finally bridging faults. The proposed parallel TRC checker has been designed in the standard $0.18\mu m$ CMOS technology of ST Microelectronics for a variety of n-variable values (number of inputs) ranging from 8 to 512 and the operation has been verified by electrical simulations in a full range of PVT (Process, Voltage, Temperature) conditions, that is: a) the process corners for the used technology provided by ST, b) power supply variations up to 10% and c) temperature variations from 0^oC to 125^oC. In Table 1 design issues and simulation results are presented for the proposed checker and the checker presented in [5].

According to Table 1, the proposed in this work checker is superior over the checker in [5] with respect to the required silicon area and the response delay time, especially for high values of the n-variable.

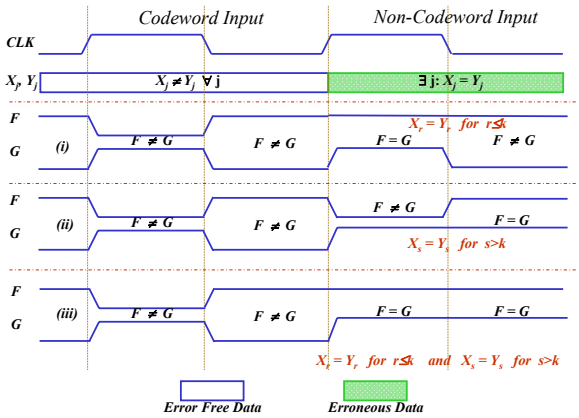


Fig. 3. Checker's response under codeword and non-codeword inputs

Table 1. Comparisons with respect to i) silicon area, ii) response delay time and iii) power consumption

Fan-in	Silicon Area Cost (UST)			Response Delay (ps)			Power Consumption (μW)		
-n-	Proposed	[5]	Reduction	Proposed	[5]	Reduction	Proposed	[5]	Reduction
8	187	252	25.8%	298	355	16.1%	64	21.3	-200.5%
16	239	415	42.4%	328	505	35.0%	95	27.7	-243,0%
32	342	776	55.9%	375	775	51.6%	109	40	-172.5%
64	540	2324	76.8%	408	1205	66.1%	119	87.1	-36.6%
128	924	7925	88.3%	517	1845	72.0%	136	245	44.5%
256	1663	34965	95.2%	695	3097	77.6%	174	976	82.2%
512	3106	135675	97.7%	1055	4965	78.8%	215	3840	94.4%

In order to extend the self-checking property of the circuit to the uncovered TSOP faults a modified version is presented in Fig. 4. In the new circuit there is a fifth nMOS transistor in the group of the four transistors (Fig. 2) that is controlled by a select signal S_j. The select signals S_j ($j \in [1, \ldots, n]$) are generated by a Cyclic Shift Register (CSR) of $k = n/2$ bits and a NOR gate array [11]. The S_j signals get successively one after the other the value "1" and thus test for TSOP the four transistors of the group including the fifth transistor that is driven by the signal S_j. It is proved [11] that the modified checker satisfies the self-checking property with respect to the same set of faults as in its previous version including the TSOP faults for the input transistors, in case that this is imperative for the design [8, 9]. Note that the parallel TRC checker presented earlier in [5] does not provide a full coverage of the TSOP faults. The proposed checker needs the application of only two codewords to satisfy the TSC or SCD properties, similarly to the checkers in [4, 5]. This is a very important property for embedded checkers.

The modified version of the proposed parallel two-rail code checker has been also designed in the same $0.18\mu m$ CMOS technology ($V_{DD} = 1.8V$), for n-variable ranging from 8 to 512. The operation of the checker has been verified by electrical simulations in a full range of PVT conditions, for all possible con-

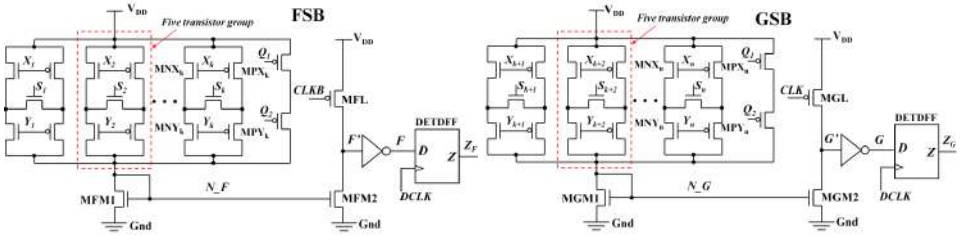


Fig. 4. The modified two-rail code checker, $k = n/2$

ditions. Monte Carlo mismatch analysis has been performed and the correct operation has been verified.

3 A Circuit for Concurrent detection of Soft and Timing errors in Digital CMOS ICs

The second analogue technique is a new soft and timing error detection circuit. It exploits the time redundancy approach that has been adopted in recent works [12, 13] and provides error tolerance in case that it will be combined with a retry cycle; that is, the correct result is obtained, each time an error is detected, by repeating the last operation using a lower frequency.

Fig. 5 presents a Functional Circuit consisting of the combinational part and the Flip-Flops of the output register. Transient faults on internal nodes of the combinational circuit may result in the appearance of transient pulses at its output lines OUT. In case that the triggering edge of the clock CLK arrives just after the transient pulse appearance and during its presence on the $OUT_{(a)}$ line (time interval δ), a soft error is generated at the output FFO of the Flip-Flop.

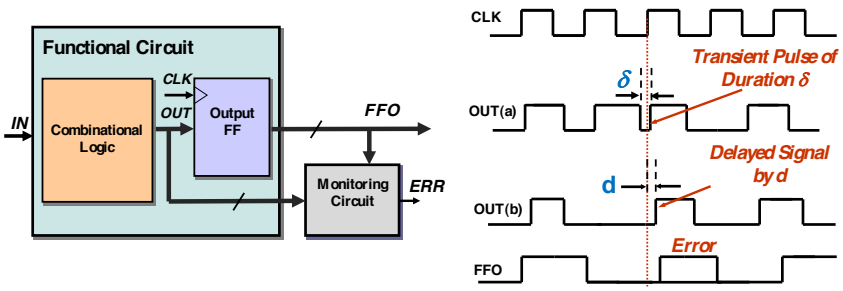


Fig. 5. Error generation mechanisms and error detection using a monitoring circuit

Moreover, path delay faults in the combinational circuit may result in a delayed signal arrival at a circuit output $OUT_{(b)}$, after the triggering edge of the

clock CLK (time interval d) and thus the generation of a timing error at the output FFO of the Flip-Flop. The key idea behind the adopted error detection technique is the use of a Monitoring Circuit to monitor the responses at the outputs of the Combinational Logic and the whole Functional Circuit after a time interval T from the latching edge of the clock signal CLK [12, 13].

In the fault free case no signal transitions appear on the monitored lines after the latching edge of the clock signal CLK and the error indication signal of the Monitoring Circuit (ERR) remains "low". In the case that a transient or a delay fault in the combinational logic causes a transient pulse or a delayed signal response (transition) on the output line OUT of the Combinational Circuit when the latter is sampled by the clock CLK, the Output Flip-Flop captures an erroneous value and an error occurs on its output FFO. The Monitoring Circuit detects the resulted difference between the values on the lines OUT and FFO and the error indication signal (ERR) rises to "high".

The proposed Monitoring Circuit that exploits a sense amplifier for soft and timing error detection is shown in Fig. 6 and consists of a Pre-Sensing Block (PSB), a Sense Amplifier (SA) and an Error Indication Flip-Flop (EIFF). The Pre-Sensing Block is divided into two sub-blocks (SBL and SBR) each one feeding a separate input of the sense amplifier INL and INR respectively.

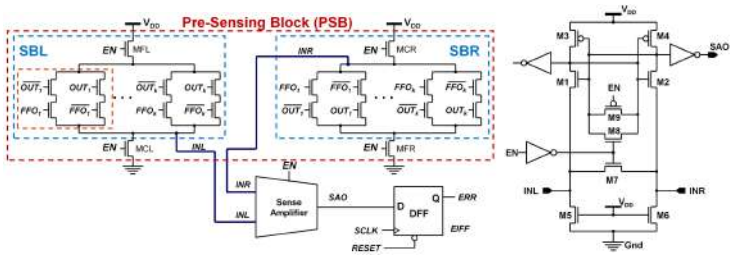


Fig. 6. The proposed Monitorting Circuit (left) and the Sense Amplifier (right)

The k pairs of monitored lines OUT_j and FFO_j ($j \in [1 \ldots k]$), are driving both sub-blocks of the Pre-Sensing Block. The SA is activated by the EN signal and provides the output signal SAO, which is latched by the Error Indication Flip-Flop (EIFF). During the system operation each period of the clock CLK can be seen as divided in two phases, the normal phase and the monitoring phase, which are defined by the EN signal, as it is shown in Fig. 7. In the normal phase, the Monitoring Circuit is inactive (EN="low").

In the monitoring phase EN="high". In the error free case where $OUT_j = FFO_j (\forall j \in [1 \ldots k])$ the SA will amplify the signal difference between its two inputs driving fast its output SAO to "low". In the presence of an error the SA will also amplify the signal difference between its inputs driving fast its output SAO to "high" providing the indication of error detection.

The $0.18\mu m$ CMOS technology of ST Microelectronics with $1.8V$ power supply has been exploited for the design of the proposed error Monitoring Circuit.

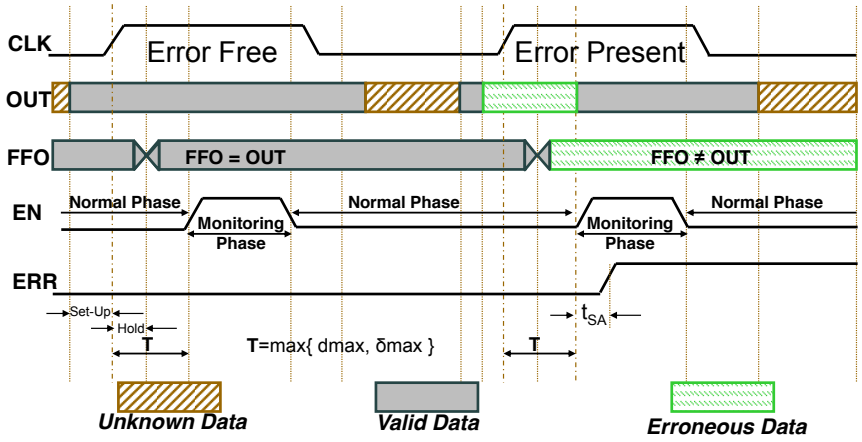


Fig. 7. Signals' timing for the Monitoring Circuit

The case of 72 monitored pairs and the corresponding layout design is given in Fig. 8. The "folded bit-line" design technique, is exploited in order to achieve a high density PSB and make the Monitoring Circuit insensitive to process and temperature variations.

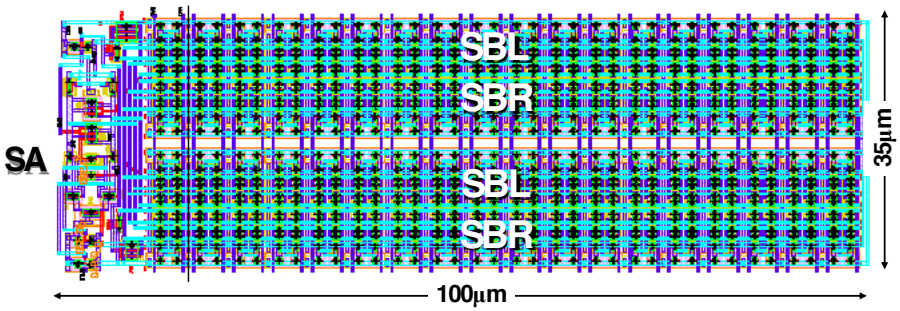


Fig. 8. Layout of the Monitoring Circuit for 72 monitored pairs of lines

Simulations and Monte Carlo analysis have been carried out [14, 15] for various numbers of monitored pairs (from 9 to 576), for temperatures up to 125^oC and using all process corner conditions. Table 2 presents comparisons between the detection times reported in [13] and the corresponding times in this circuit for various numbers of monitored pairs [16]. These measurements have been carried out at 125^oC and for the slow-slow transistor model, which, according to the simulations, provides the worst case response times.

Table 2. Detection time comparisons

Number of Monitored Pairs	Detection time (ps)		Reduction (%)
	[13]	Proposed	
9	456	191	58
18	501	226	55
36	581	265	54
72	721	317	56
144	979	376	62
288	1485	430	71
576	2480	468	81

4 Coping with current variations in I_{DDQ} testing

The quiescent current (I_{DDQ}) of a circuit is defined as the sum of its leakage currents (background current I_B), plus any defective current (I_{DEF}). I_{DDQ} monitoring is a well established technique for testing integrated circuits (ICs) in CMOS technologies. I_{DDQ} testing is based on the assumption that the intrinsic, defect-free, quiescent current of an IC is small compared to the quiescent current in the presence of a defect in the circuit. Consequently, setting the maximum current from the expected range of background currents in a circuit under test (CUT) as the threshold current, we can discriminate defect free from defective ICs by comparing their I_{DDQ} current with this threshold current.

Fig. 9 (left) presents an I_{DDQ} testing scheme based on the use of a Current Sensor (CS), either embedded to the IC (Built-In Current Sensor - BICS) or external to it. The Circuit Under Test (CUT) is isolated from the ground supply (Gnd) by MN_G transistor while the current sensor is connected to the virtual ground (V_Gnd) of the CUT. During the normal mode of operation the V_Gnd node is grounded. In the test mode of operation the signal T_ENB turns low and the CS compares the I_{DDQ} current of the CUT with a reference current (I_{REF}). In case that the I_{DDQ} current is greater than the I_{REF} current, the CUT is characterized as defective. According to the above scheme, the I_{REF} current must be greater than the maximum defect free background current I_B of the CUT.

In nanometer technologies the circuit background current I_B is increased with technology evolution [18]. Moreover, the defective current I_{DEF} that is required to be detectable is decreased [17]. In addition the number of transistors in a single chip is increased rapidly resulting in the reduction of the gap between the values of defect free and defective I_{DDQ} currents. Furthermore, the value of I_B is also influenced by temperature and increased process variations. Therefore, the application of I_{DDQ} testing using a unique reference current I_{REF} for discrimination between defect free and defective circuits for all chips in a production line, is impractical since it will either lead to yield loss or reduced fault coverage. Consequently, I_{REF} must be adjusted for each chip in order to take into account process variations.

The circuit in Fig. 9 (right) uses an extra transistor MN_T in parallel to MN_G, proper biased by voltage V_{bias} so that in the defect free case the volt-

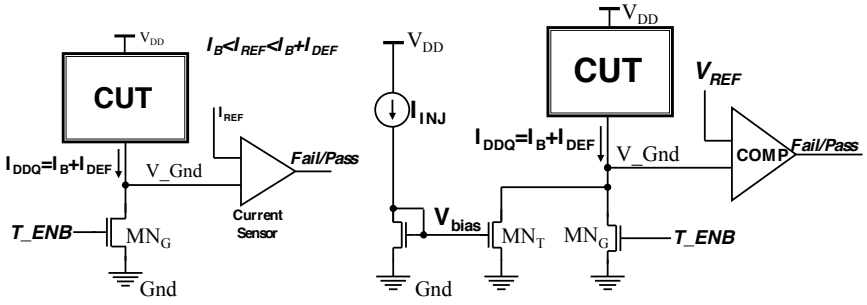


Fig. 9. A common I_{DDQ} testing scheme (left). The adjustable I_{DDQ} testing concept (right).

age at the virtual ground node (V_Gnd) is less than a reference voltage V_{REF}. The bias voltage V_{bias} can be generated using an injection current I_{INJ} and a current mirror. However, since the background current I_B of the CUT is influenced by process and temperature variations, the injection current I_{INJ} must be accordingly adjusted in order to avoid fault coverage reduction or yield loss.

In order to dynamically adjust I_{INJ} to process and temperature variations we adopted the partitioning of the CUT into two subcircuits (the left subcircuit sub-CUTL and the right subcircuit sub-CUTR). Then the background current of the left subcircuit is used as injection current (I_{INJ}) for the testing of the right subcircuit and vice-versa. Since in each case the background and the injection currents are influenced by the same process and temperature variations in the CUT, the I_{DDQ} testing process turns to be almost independent of these two factors.

In Fig. 10 the simplified block diagram of the proposed I_{DDQ} testing technique is presented, where the background current of sub-CUTL is used to generate the injection current for the I_{DDQ} testing of sub-CUTR. A preliminary study of this I_{DDQ} testing architecture and the built-in current sensing (BICS) circuit has been presented in [19] while early experimental results were discussed in [20] and [21].

The I_{DDQ} testing circuitry (consisting of the CMA, the comparator and transistors MN_{GL} and MN_{GR}) can be either embedded in the chip, forming a BICS circuit, or externally. Each partition must have a dedicated virtual ground (V_Gnd_L and V_Gnd_R respectively). In general the two subcircuits under consideration during I_{DDQ} testing are not identical. Consequently, their background currents I_{BL} and I_{BR} are not expected to be equal. In addition, the magnitude of each background current depends on the applied test vector. From the above it is evident that a tunable current mirror (a current mirror with tunable current gain β) is required in order to be able to generate for each test vector (j) the bias current $I_{B(L/R)j}$ from the injection current $I_{B(R/L)j}$ according to the following relation: $I_{B(L/R)j} = \beta_j I_{B(R/L)j}$. The proposed implemented tunable current mirror amplifier (T-CMA) is illustrated in Fig. 11.

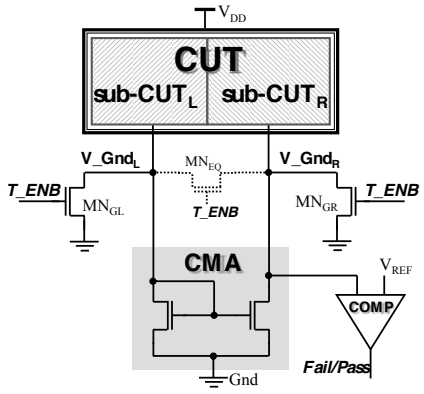


Fig. 10. The proposed I_{DDQ} testing technique

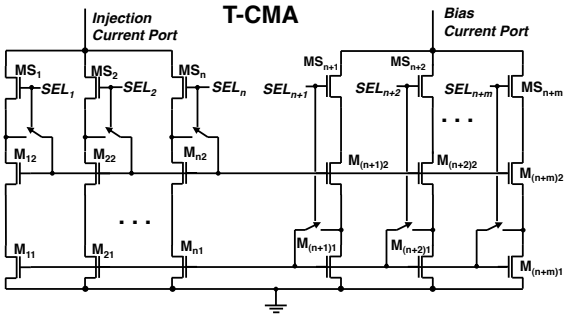


Fig. 11. A tunable current mirror amplifier (T-CMA)

In order to validate the proposed I_{DDQ} testing technique a demonstration circuit (consisting of a digital circuit and a BICS circuit) has been designed and fabricated (see Fig. 12) in a standard $180nm$ CMOS technology ($V_{DD} = 1.8V$). The digital circuit has been partitioned into two subcircuits. The microphotograph of the demonstrator is shown in Fig. 12.

Also in this Thesis a comprehensive theoretical analysis of the proposed technique is provided, in order to have a quantitative estimation of the trade-off between resolution (res), size of the partition of the CUT (N) and the size of the BICS.

The defective current resolution (res) is defined as the minimum amount of defective current that the BICS can distinguish to the total fault free background current of the CUT. In I_{DDQ} testing we want the resolution to be as small as possible so that small defective currents, or in other words high defective resistances (lighter defects), are detectable. From the analysis it is shown that as the circuit size is increased, a desired defective current resolution can be achieved by increasing the current mirror transistor widths. In Fig. 13 the defective current

Fig. 12. Fabricated I_{DDQ} test chip and a microphotograph

resolution as a function of the transistor widths (W_R) in the current mirrors is presented for various circuit sizes (N).

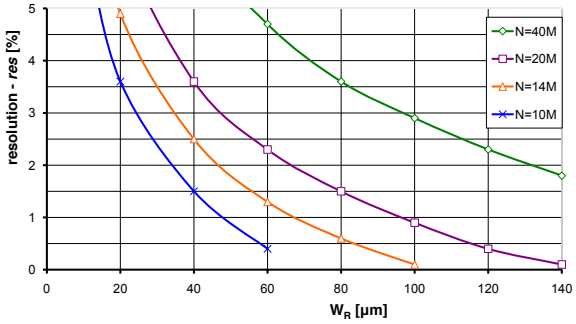


Fig. 13. Defective current resolution with respect to current mirror transistor width W_R for various circuit sizes N

The experimental results from the fabricated demonstration circuit confirmed that the proposed I_{DDQ} testing technique is capable to provide high fault coverage for the circuit under test avoiding yield loss.

References

1. D.A. Anderson and G. Metze, "Design of Totally Self-Checking Circuits for m-out-of-n Codes," *IEEE Trans. on Computers*, vol. 22, pp. 263–269, 1973.
2. M. Nicolaidis and B. Courtois, "Strongly Code-Disjoint Checkers," *IEEE Trans. on Computers*, vol. 37, pp. 751-756, 1988.
3. S.J. Piestrac, "Design Method of a Class of Embedded Combinational Self-Testing Checkers for Two-Rail Codes," *IEEE Trans. on Computers*, vol. 51, no. 2, pp.229–234, Feb. 2002.

4. S. Kundu, E.S. Sogomonyan, M. Goessel and S. Tarnick, "Self-Checking Comparator with One Periodic Output," *IEEE Trans. on Computers*, vol. 45, no. 3, pp. 379-380, 1996.
5. M. Omana, D. Rossi and C. Metra, "Low Cost and High Speed Embeded Two-Rail Code Checker," *IEEE Trans. on Computers*, vol 54, no 2, pp. 153-164, 2005.
6. S. Matakias, Y. Tsiatouhas, Th. Haniotakis A. Arapoyanni, and A.Efthymiou, "Fast, Parallel Two-Rail Code Checker with Enhanced Testability," in *11th IEEE International On-Line Testing Symposium (IOLTS)* 2005, pp 149–156.
7. International Technology Roadmap for Semiconductors, `http://public.itrs.net/`.
8. R.R. Montanes, P. Volf and J.P. de Gyvez, "Resistance Characterization for Weak Open Defects," *IEEE Design and Test of Computers*, vol. 19, no. 5, pp. 18-26, Sept./Oct. 2002.
9. J. Jahangiri and D. Abercrombie, "Value-Added Defect Testing Techniques," *IEEE Design and Test of Computers*, vol. 22, no. 3, pp. 224–231,May/June 2005.
10. S. Matakias, Y. Tsiatouhas, Th. Haniotakis and A. Arapoyanni, "Ultra Fast and Low Cost Parallel Two-Rail Code Checker Targeting High Fan-In Applications," in *IEEE Computer society Annual Symposium on (ISVLSI)*, pp. 293–296, 19-20 February 2004.
11. S. Matakias, Y. Tsiatouhas, Th. Haniotakis, A. Arapoyanni, "A Current Mode, Parallel, Two-Rail Code Checker," *IEEE Trans. On Computers*, vol. 57, No. 8, pp 1032–1045, August 2008.
12. L. Anghel and M. Nicolaidis, "Cost Reduction and Evaluation of Temporary Faults Detecting Technique," *Design Automation & Test in Europe*, pp. 591–598, 2000.
13. Y. Tsiatouhas, A. Arapoyanni, D. Nikolos and Th. Haniotakis, "A Hierarchical Architecture for Concurrent Soft Error Detection Based on Current Sensing," in *8th IEEE Int. On-Line Testing Workshop*, pp. 56–60, 2002.
14. Y. Tsiatouhas, S. Matakias, A. Arapoyanni and Th. Haniotakis, "A Sense Amplifier Based Circuit for Concurrent Detection of Soft and Timing Errors in CMOS ICs," in *9th IEEE International On-Line Testing Symposium (IOLTS)*, pp 12–16, 7-9 July 2003.
15. S. Matakias, Y. Tsiatouhas, A. Arapoyanni and Th. Haniotakis, "A Circuit for Concurrent Detection of Soft and Timing Errors in Digital CMOS ICs," *Special Issue of Journal of Electronic Testing: Theory and Applications*, vol. 20, pp 523–531, 2004.
16. S. Matakias, Y. Tsiatouhas, A. Arapoyanni, Th. Haniotakis, "A High Speed Circuit for Concurrent Detection of Soft Errors in CMOS ICs," *Radiation Effects on Circuits and Systems (RADECS)*, pp A8 1-4, 2006.
17. R.R. Montanes and J. Figueras, "Estimation of the Defective I_{DDQ} Caused by Shorts in Deep Submicron CMOS ICs," in *Design Automation and Test in Europe (DATE)*, pp. 490–494, 1998.
18. S. Henzler, "Power Management of digital Circuits in Deep Sub-Micron CMOS Technologies," Springer, 2007.
19. Y. Tsiatouhas, Th. Haniotakis, D. Nikolos and A. Arapoyianni, "Extending the Viability of I_{DDQ} Testing in the Deep Submicron Era," in *IEEE International Symposium on Quality Electronic Design (ISQED)*, pp. 100–105, 2002.
20. S. Matakias, Y. Tsiatouhas, A. Arapoyanni, Th. Haniotakis, G. Prenat and S. Mir, "A Built-In I_{DDQ} Testing Circuit," in *31st European Solid-State Circuits Conference (ESSCIRC)*, pp 471–474, 12-16 September 2005.
21. S. Matakias, Y. Tsiatouhas, A. Arapoyanni, Th. Haniotakis, "An Embedded I_{DDQ} Testing Circuitand Technique," in *12th IEEE International Conference on Electronics, Circuits and Systems*, 11-14 December 2005.

Investigation of optical wireless systems for indoor broadband networks

Georgia Ntogari*

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
gntogari@di.uoa.gr

Abstract. Today's home area networks (HANs) are designed to satisfy the expectations of subscribers for access in bandwidth hungry services, such as High Definition TV - Video, Telepresence, 3D Gaming, Virtual Reality and e-Health. To this direction the use of wireless optical systems is examined. A simulation model was developed in MATLAB to compute the impulse response of the optical wireless channel and determine the characteristics and the restrictions imposed by each channel topology. Appropriate equalization and multiple input multiple output (MIMO) techniques were investigated as a means of fighting intersymbol interference and increasing the coverage area and data rate. In an effort to increase the receiver's sensitivity under the intense ambient light noise the use of coherent sources and homodyne detection is proposed. It is shown that coherent optical wireless systems in combination with space-time diversity enable the transmission of data with rates in the order of ~Gb/s. Finally, the implications of the incorporation of a dimming functionality in the light emitting diodes (LEDs) of a visible light communication (VLC) system are investigated.

Keywords: optical wireless networks, coherent detection, equalization, space time block coding, dimming

I. Introduction

As the demand for ultra broadband wireless access home networks constantly increases, the radio frequency spectrum is becoming extremely congested and thus, attention is drawn towards alternative technologies. Indoor infrared wireless communications were first proposed by Gfeller and Bapst [1] and are since attracting growing interest due to the abundance of unregulated bandwidth, which renders them an attractive candidate for high speed data communications. The infrared channel is not without drawbacks, however. In many indoor environments, it is not easy to achieve a high Signal-to-Noise (SNR) ratio, since there may be intense ambient infrared noise [2]. This noise is due to the infrared spectrum components arising from the radiation of tungsten or fluorescent lamps and sunlight. In addition, artificial light introduces significant in-band components for systems operating at bit rates up to several Mb/s and thus induces interference [3], [4]. Moreover, the power constraints on infrared transmitters imposed by eye-safety regulations, may limit the range of these systems. Infrared links are also susceptible to shadowing caused by objects or people positioned between the transmitter and the receiver.

The effect of shadowing can be dealt with, by using non directed configurations, in which the optical link does not depend on the Line Of Sight (LOS) path between the transmitter and the receiver. Compared to LOS systems, non directed configurations suffer from higher path loss and require higher levels of transmitted power and larger photodetecting area at the receiver. The multipath propagation observed, gives rise to intersymbol interference (ISI), which becomes critical at high data rates. Nevertheless, to date, the non directed configurations, have received great interest from the research community, and a number of experimental links has been reported covering bit rates up to 50 Mb/s [5].

Optical wireless communications with (light emitting diodes (LEDs) emitting in the visible spectral range has recently gained increasing attention and is commonly referred to as visible-light communications (VLC) [6]-[9]. In some use cases, LEDs in VLC systems serve a dual role in providing both illumination and wireless connectivity [8]. The

* Dissertation advisor: Thomas Sphicopoulos, Professor.

pioneering idea of using white LEDs lighting systems for communication is attributed to the research group of Nakagawa [8]. Phosphorescent white LEDs have a limited modulation bandwidth (~ 2MHz), however, by placing a blue optical filter in front of the receiver, the modulation bandwidth can be extended to 20MHz [10]. Using discrete multitone (DMT) modulation [11], VLC systems can provide >200Mb/s transmission rates with commercial high-power lighting LEDs [12]. Dimming is an essential functionality of modern lighting systems. In the case of LEDs, pulse-width modulation (PWM) seems to constitute the most effective means of accurately controlling LED illumination without incurring color rendering of the emitted light [13], [14].

The objective of this work is to examine the capacity limits of non directed indoor infrared wireless systems assuming different transmitter and receiver configurations in combination with appropriate equalization and multiple input multiple output (MIMO) techniques. In addition, the potential of MIMO coherent optical wireless systems is examined. Finally, the possibility of combining PWM dimming with VLC, based on DMT, on the physical layer is considered both analytically and numerically.

II. Results and Discussion

In order to evaluate the effect of different transmitter-receiver configurations on the performance of a wireless infrared link, a tool that computes the impulse response of the infrared channel for each configuration, based on the modified Monte Carlo method [15], was developed in MATLAB. For the simulations a medium-sized office room, depicted in the inset of Fig. 1 was considered. Table I, outlines the basic configuration parameters for the simulation. In the table, ρ_{north}, ρ_{south}, ρ_{east}, ρ_{window}, $\rho_{ceiling}$ and ρ_{floor} denote the reflectivities of the corresponding surfaces of the room, L_x, L_y and L_z are the room dimensions along the x, y and z axis respectively, depicted in the inset of Fig. 1 and HPSA is the half power semi angle of the transmitter, which is related to the order m of the transmitter radiation pattern through m=-ln2/ln(cos(HPSA)). Two transceiver configurations were considered [16]. In the first configuration, classified as vertically oriented, i.e. T1R1, the main lobe of the transmitter and the receiver is directed upwards, towards the ceiling. In the second one, classified as horizontally oriented, i.e. T8R8, some of the lobes are also directed parallel to the ceiling, potentially offering a LOS path and possibly higher coverage.

Table I
Configuration Parameters

PARAMETERS	*T1R1*	*T8R8*
Room		
(L_x,L_y,L_z)	(5.5,7.5,3.5)	(5.5,7.5,3.5)
ρ_{east}	0.3	0.3
ρ_{south}	0.56	0.56
ρ_{north}	0.3	0.3
ρ_{window}	0.04	0.04
$\rho_{ceiling}$	0.69	0.69
ρ_{floor}	0.09	0.09
Transmitter		
HPSA	1x60°	6 x 30°+2 x 30°
Azimuthal separation	0	6 x 45°
elevation	1 x 90°	6 x 0°+2 x 90°
position	(2,4,1.5)	(2,4,1.5)
Receiver		
FOV(φc)	60°	31°
Position	NW-SE diagonal height: 0.8m	NW-SE diagonal height: 0.8

For the T1R1 configuration, the transmitter has a first order Lambertian pattern and is oriented vertically towards the ceiling. The receiver is a pin photodetector of area A_{det}=1cm^2 with an optical concentrator having cutoff angle of 60° and refractive index n_c=1.44, while the optical filter has a bandwidth $\Delta\lambda$=50nm. For the T8R8 configuration the transmitter uses six equal power 30° HPSA transmit beams equally spaced in the horizontal plane and two such identical beams pointing straight up. The receiver uses eight optical concentrators with cut-off angles 31°, seven of which are horizontally oriented and one is pointing straight up. The power collected from each receiver is added together to obtain the total

received power. The transmitter has a center wavelength of 806nm and is located at a height of 1.5 m, near the center of the room. In order to incorporate the effect of the ambient light noise in the room under investigation, eight incandescent lamps were considered at the ceiling and the west wall of the room was assumed to be a window (in order to account for the sunlight) as is depicted in the inset of Fig. 1.

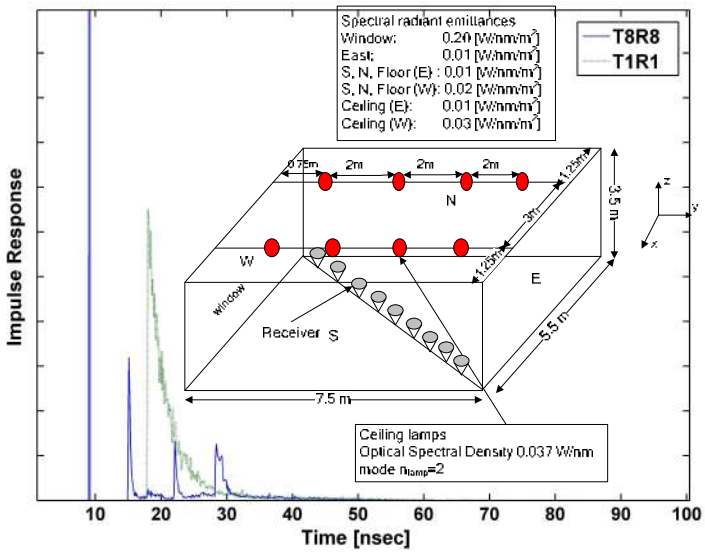


Fig. 1. Impulse response of the optical wireless channel for configurations T8R8 and T1R1.

Taking into account the previously described system parameters, the impulse response was calculated for all the receiver positions along the south-east north-west diagonal of the room and it is depicted in Fig. 1 for the two configurations T1R1 and T8R8. It is deduced that both channels introduce time dispersion, which for high data rates can lead to severe ISI and hence to performance degradation. The T8R8 configuration is bound to induce higher ISI for bit rates >100Mb/s than T1R1, since the corresponding time dispersion is >20nsec. To this end, the employment of appropriate equalization schemes, i.e. linear (LE) and decision feedback equalization (DFE) was proposed as well as the use of multicarrier schemes, i.e. DMT with MIMO.

A. Equalization schemes for diffuse IR systems

In an effort to mitigate the effects of ISI, several detection schemes have been proposed [5]. In the case of the unequalized system, the SNR is given by [17]:

$$SNR_U = \min_{(i,j)} \left\{ \left\langle \frac{\left(m_i - m_j\right)^2}{2N_0} \right\rangle \right\} \tag{1}$$

where m_i is the received signal power when symbol i is transmitted and N_0 the ambient light noise power. In the presence of ISI, for a symbol transmitted at time t_0=0, one needs to calculate the values of SNR_U considering the adjacent symbols at $\pm kT$, $k\neq 0$. The parameters m_i are calculated using:

$$m_i = P_p \sum_k a_k \left(\frac{1}{T} \int_{-T/2}^{T/2} p(\tau - kT) d\tau \right) \tag{2}$$

where, P_p the peak power, L is the number of the different transmitted symbols a_k according to the modulation scheme of choice (L=2 for OOK and L=4 for 4-PPM and 4-PAM) and $T=\log_2(L)/R_b$ is the symbol duration while R_b is the bit rate and assuming that the values of the symbol sequence a_k are such that the symbol transmitted at t_0=0 corresponds to i. In (2), $p(t)$ is a rectangular pulse $rect(t)$ (height=1 and width=T) passed through a baseband filter which represents the combined effects of the transmitter shaping, the infrared channel

propagation and the photodiode responsivity. The values obtained by (2) are averaged with respect to the adjacent symbols at $\pm kT$, $k \neq 0$.

The optimum system performance in terms of SNR is obtained when the receive filter is matched to $p(t)$ and is defined as the Matched Filter Bound (MFB) given by [17]:

$$SNR_{MFB} = \frac{P_p}{N_0} \int_{-\infty}^{+\infty} M(f) df \tag{3}$$

where

$$M(f) = \frac{1}{S_n(f)} \left| \int_{-\infty}^{+\infty} p(t) e^{j2\pi ft} \right|^2 \tag{4}$$

In (3), $M(f)$ is the frequency spectrum of the matched filter's output pulse. Alternatively, LE or DFE equalization schemes are suboptimal strategies for detecting signals in the presence of ISI, their primary advantage being a reduction in complexity. For the LE equalizer, the SNR is given by [17]:

$$SNR_{LE} = \frac{P_p}{N_0} \left(\int_{-1/2T}^{1/2T} \frac{df}{S(f)} \right)^{-1} \tag{5}$$

while for the DFE, the SNR becomes [17]:

$$SNR_{DFE} = \frac{P_p T}{N_0} \exp\left(T \int_{-1/2T}^{1/2T} \ln\left[S(f) \right] df \right) \tag{6}$$

The spectrum $S(f)$ is given by:

$$S(f) = \frac{N_0}{P_p T} + \frac{1}{T^2} \sum_k M\left(f + \frac{k}{T} \right) \tag{7}$$

The transmit power equals 0.6W and the bit-rate of the system under examination is 100 Mbps. Three modulation schemes were investigated, 4-PPM (pulse position modulation), OOK (on-off keying) and 4-PAM (pulse amplitude modulation). The electrical SNR obtained for the configurations T1R1 and T8R8 is depicted in Fig. 2 and Fig. 3 respectively, for 4-PAM which outperforms OOK and 4-PPM [18], [19], when different equalization schemes are employed.

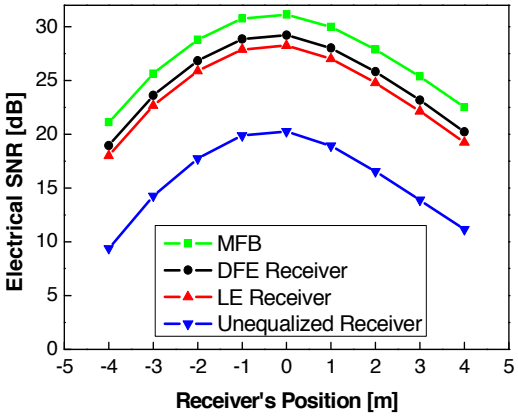


Fig. 2. SNR for T1R1 when 4-PAM is employed.

In Fig. 2 according to the MFB curve, the SNR cannot exceed the value of 30 dB at the center of the room whereas near the corners it does not drop below 20 dB. The DFE and LE schemes improve the performance of the unequalized system by almost 9 and 8 dB respectively.

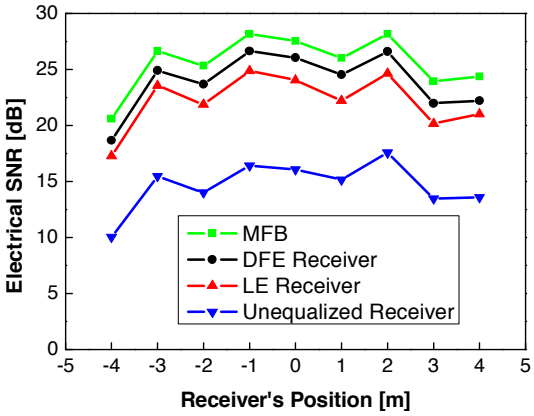


Fig. 3. SNR for T8R8 when 4-PAM is employed.

Better coverage can be obtained using the T8R8 transmitter/receiver configuration. The values of the SNR obtained at different receiver positions, are depicted in Fig. 3. Comparing these values with the ones in Fig. 2 it is deduced that there are no large variations in the values of the SNR and hence, the system performance is not expected to vary significantly (except at the edges of the room).

As in the case of T1R1, both the LE and DFE equalization techniques significantly improve the system performance. For example, if one excludes the SNR values obtained at receiver positions near the two edges of the diagonal, the SNR for T8R8 is higher than 20dB implying a BER much less than 10^{-14}. The worst case SNR is again obtained at the edge of the room and for the case of the equalized schemes is approximately the same as those obtained by T1R1. The variations in the SNR values at different positions along the main diagonal of the room can be interpreted in combination to the impulse response obtained for both configurations, see Fig. 1. It is deduced that in the T8R8 impulse response four peaks are observed while in the T1R1 only one. This can be attributed to the horizontal transmit and receive lobes of the T8R8 configuration and it is the reason for the different shapes of the SNR distribution between T8R8 and T1R1. These results seem to indicate that the T8R8-4-PAM configuration can carry ≥100Mb/s (Fast Ethernet type) data rates in almost every point in the room and should be considered favorably as a potential hot spot for future indoor WLANs.

B. Diversity schemes for diffuse infrared wireless systems

This section provides a framework for the performance evaluation of an N MIMO DMT M-QAM (quadrature amplitude modulation) system. It will be shown that using multiple transmitter arrangements in diffuse optical wireless is interesting, since the power requirements per transmitter can be reduced, while system performance can be enhanced. Alamouti-type Space Time Block Coding (STBC) [20] and repetition coding is investigated and compared to Single Input Single Output (SISO) and Maximum Ratio Combining (MRC) [21] taking into account the channel's frequency response as well as the noise distribution throughout the simulation room. In order to mitigate the effects of the channel's frequency response, which is necessary for STBC, DMT modulation with QAM is used.

Assuming Q transmitters the optical power of the q^{th} transmitter is written as:

$$P^q(t) = \frac{A}{Q}\mathrm{Re}\left\{\sum_{k=-\infty}^{+\infty}\sum_{n=1}^{N-1} s_{k,n}^q e^{j2\pi f_n t} p(t-kT)\right\} + \frac{P_{DC}}{Q} \tag{8}$$

where A is the amplitude of the DMT waveform, $p(t)$ is the QAM pulse shape, $s_{k,n}$ is the k^{th} QAM symbol transmitted at the n^{th} subcarrier channel centered around $f_n=n/T$, T is the DMT signal duration and P_{DC}, is a DC power level added to ensure that the optical power is always positive. The symbols $s_{k,n}=a_{k,n}+jb_{k,n}$ take their values from a QAM constellation. Hence, the possible values of $a_{k,n}$ and $b_{k,n}$ are given by $2u-(M^{1/2}+1)$ where $1\le u\le M^{1/2}$ and M is the number of QAM distinct symbols in the constellation assumed to be a power of 2, i.e. $M=2^L$. Usually L is an even number. The duration of the QAM symbol is given by $T=\log_2(M)/R_b$ where R_b is the bit rate. The pulses $p(t)$ are assumed ideal rectangular pulses, such that $p(t)=1$ for $0\le t\le T$ and $p(t)=0$ otherwise. Note that equation (8) assumes that the DC

subcarrier at f_0=0 is not modulated, as in [22] in order to avoid interference problems from electronic ballasts.

In the absence of clipping and since $P(t)\geq 0$ one can choose the amplitude A such that $\min\{P(t)\}\geq 0$. Taking into account that:

$$\left|\mathrm{Re}\left\{\sum_{n=1}^{N-1} s_{k,n} e^{j2\pi f_n t}\right\}\right| \leq \left|\sum_{n=1}^{N-1} s_{k,n} e^{j2\pi f_n t}\right| \leq \sum_{n=1}^{N-1} \left|s_{k,n}\right| \tag{9}$$

and that $\max\{|s_{k,n}|\}=2^{1/2}[(M)^{1/2}\ 1]$, it is easy to show that the following choice for A,

$$A = \frac{P_{DC}}{\sqrt{2}\left(\sqrt{M}-1\right)\left(N-1\right)} \tag{10}$$

At the receiver the SNR for the m^{th} subcarrier is computed by:

$$SNR_m = \left(RA/2Q\right)^2 V_m / N_0 \tag{11}$$

where R is the receiver's responsivity factor, N_0 the noise power and V_m the channel matrix which depends on the selected scheme. For Alamouti STBCs V_{m} is $V_m=\sum_{p,q}|h_{p,q}(\mathrm{f_m})|^2$ while for repetition coding $V_m=(\sum_{p,q}|h_{p,q}(\mathrm{f_m})|)^2$ where $h_{pq}(f)$ is the channel transfer function between the p^{th} receiver and the q^{th} transmitter[2]. The Bit Error Rate (BER) is computed by [23]:

$$P_e = \frac{4}{N}\sum_{n=1}^{N-1}\frac{\sqrt{M}-1}{\sqrt{M}\,m} Q\left(\sqrt{\frac{3mM}{M-1}SNR_n}\right) \tag{12}$$

where $m=\log_2 N$.

The configuration used in the simulations is T1R1 for the room depicted at the inset of Fig.1. The noise power distribution was computed assuming irradiance of eight incandescent lamps and sunlight. The BER was computed for various receivers' positions along the main diagonal of the room (NW to SE) at a height of 0.8 m from the floor. The transmitters are positioned along the main diagonal at equal distances from the center and 1.0 m above the floor. For the single input schemes (SISO and MRC) one transmitter was assumed, positioned in the center of the room again at a height of 1.0 m.

Three MIMO schemes were considered [24]; the first employing two transmitters and one receiver (2×1), the second two transmitters and two receivers (2×2) using repetition coding and STBC, whereas the third one transmitter and two receivers (MRC). These schemes are compared to a SISO system. In multiple receiver arrangements, the spacing between the receivers was equal to 35 cm. The results were obtained for a 100 Mb/s, 16-QAM/DMT of 32 subcarriers, with P_{DC}=0.6W.

Fig. 4 shows the values of the BER as a function of the distance d from corner A of the room (NW corner). As seen by the figure, the 2×1 Alamouti and the SISO schemes have a large BER at small values of d, in which case the receiver is situated near the window of the room (see Fig. 1) and hence the power of the ambient light noise is high. Regarding the multiple receiver systems, one notices that MRC is generally better than the 2×1 and SISO, a fact that is observed in radio wireless systems as well [20]. On the other hand, both the repetition coding and the Alamouti 2×2 systems outperform the MRC, leading to a significant performance enhancement.

It is interesting to observe that the 2×2 and 2×1 systems achieve their optimum performance near $d\approx$7.5m while the minimum BER for single transmitter schemes is obtained near the center of the diagonal of the room. This can be explained since the SNR is not the same in the four systems under consideration. In contrast to SISO and 2×1, in the case of MRC and 2×2 systems there are two receivers, each one adding a slightly different noise component. The figures also illustrate that MIMO schemes do not always guarantee a performance enhancement especially if STBC is used. Interestingly enough however, repetition coding compares more favourably to MRC as illustrated by Fig. 4. In addition repetition coding outperforms STBC and this is not surprising since the systems considered use direct detection. However, multiple transmitter arrangements have the advantage of using lower power transmitters, relaxing the power constraints due to eye or skin safety.

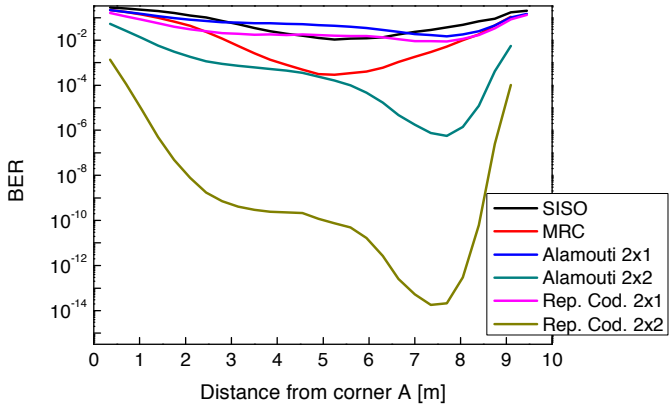


Fig. 4. Comparison of SISO, MRC, Alamouti 2×1 and 2×2, and Repetition Coding 2×1 and 2×2 schemes for various receiver positions along the main diagonal of the room.

C. Diversity schemes for coherent LOS infrared wireless systems

Inspired by the fact that unlike direct detection, in coherent detection the photocurrent is proportional to the incident electric field we will examine in this section the potential of STBC schemes as a means of further improving the coverage and data rate of coherent indoor optical LOS links. To this end, analytic formulas for the BER are derived for two Alamouti schemes and are compared to a SISO system. The BER performance of the system is realistically examined taking into account both thermal and shot noise, although ideal coherent receivers are usually considered shot noise limited for a sufficient local oscillator (LO) power. It is shown that Alamouti STBC may increase the capacity of coherent LOS IR systems, improve their coverage and decrease the required optical power at the transmitter.

The SISO coherent IR system uses of a phase-shift-keying (PSK) modulator which imprints the information on the phase of the optical carrier wave emitted by the laser source. The emitted light is collimated and uniformly illuminates the coverage area through a holographic diffuser [25]. At the receiver homodyne detection is employed using a 90° optical hybrid followed by two balanced photodetectors [26]. The received and LO fields are considered linearly polarized in orthogonal directions.

The approximate closed form formula for the BER in the case of M-level PSK modulation and a 2×1 Alamouti scheme is given by [23]:

$$P_e^{2x1} = \frac{2}{\log_2 M} Q\left(\sqrt{2\sin^2\left(\frac{\pi}{M}\right)\log_2 M \frac{\|\mathbf{H}\|^2}{\sigma^2}}\right) \tag{13}$$

where $Q(x)$ is the Q function and σ^2 is the total noise power calculated by multiplying the total noise PSD given by with the bit rate, R_b, of the system and $\mathbf{H}$ is a 2×2 array with channel coefficients given by [27]:

$$H_{pl} = 2R\sqrt{\frac{P_{LO}P_T A_{eff}^{(pl)}}{2A_{\text{cov}}}} e^{j(kd_{pl}-\phi_{LO})} \tag{14}$$

where R is the receiver's responsivity, P_{LO} and φ_{LO} is the optical power and phase of the LO at the receiver, P_{T} is the average transmitted optical power, $A_{eff}^{(pl)}$ the effective receiver area, A_{cov} the total coverage area of the system and d_{pl} the distance between transmitter p and receiver l. The distance between the receivers for the 2×2 system was taken equal to 0.07m, the total transmitted optical power is P_{T}=20mW and the LO power is P_{LO}=5mW.

Following a similar approach, it is also possible to derive the BER in the case of a repetition coding 2×1 scheme. It can easily be shown that formula still holds if one replaces $\|\mathbf{H}\|=|H_{11}|^2+|H_{21}|^2$ by $|H_{11}+H_{21}|^2$. Using, it is possible to show that:

$$\left|H_{11}+H_{21}\right|^2 = \left|H_{11}\right|^2 + \left|H_{21}\right|^2 + 4R^2\frac{P_{LO}P_T A_{eff}}{A_{\text{cov}}}\cos\left(k\left(d_{11}-d_{21}\right)\right) \tag{15}$$

Equation (15) suggests that, since d_{11} and d_{21} are much larger than the wavelength $\lambda=c/f_0$, then the error probability in may undergo large fluctuations even for small changes in the terminal position

The system under investigation is located within a room of dimensions 6 × 6 × 3 m like in [25]. Using the equations outlined previously the BER for all three configurations was obtained at multiple receiver positions. A BPSK scheme was considered and the bit rate is taken at R_b=1Gb/s. The transmitters employed in the STBC schemes emit half the power than the transmitter employed in SISO.The dominant noise source in the system under study is the LO shot noise with a PSD of ≅53 pA√Hz whereas the PSD of the ambient light noise and the thermal noises calculated in the same way as in [24] turn out to be ≅0.005pA√Hz and ≅5pA√Hz respectively.

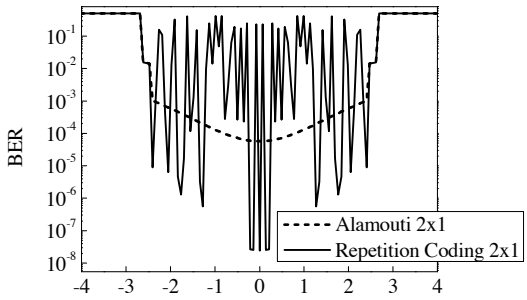


Fig. 5. BER along the main diagonal on the receiver plane for an Alamouti and a repetition coding 2×1 scheme.

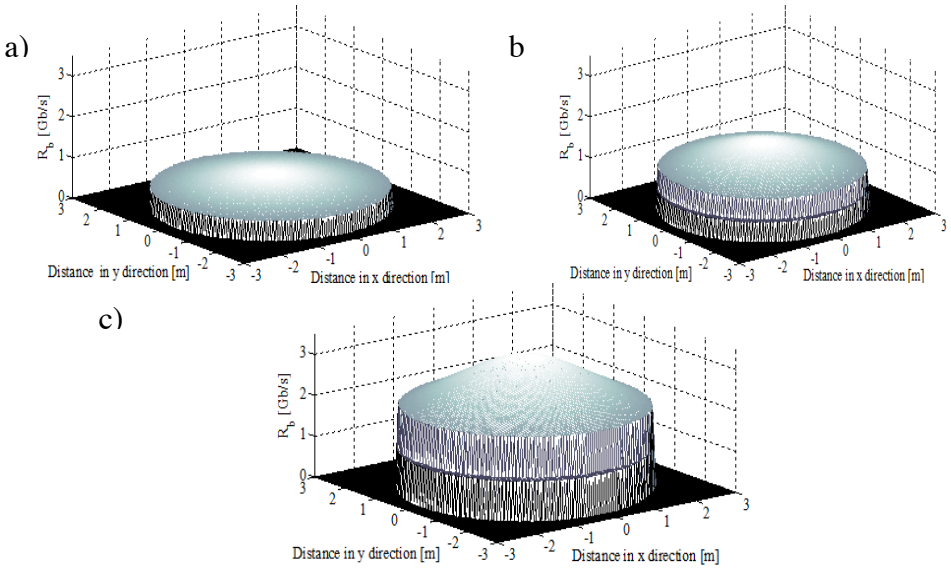


Fig. 6. Maximum R_b value throughout the room for the a) SISO scheme, b) 2×1 Alamouti STBC and c) 2×2 Alamouti STBC when a BER of 10^{-3} is assumed.

In Fig. 5 we show the BER computed along the main diagonal on the receiver plane for Alamouti and repetition coding 2×1. It is interesting to observe that while for some receiver positions the BER is lower (~10^{-7}) than the one obtained using the Alamouti scheme, for other positions which are in close proximity, the BER is extremely high (~10^{-1}) rendering communication impossible. These variations are attributed to the cosine term appearing in (15). The obtained results demonstrate that a coherent optical wireless system employing repetition coding may not provide reliable communication due to these large BER fluctuations.

The SISO, 2×1 and 2×2 STBC systems were also evaluated in terms of the achievable bit rate. To this end, the maximum bit rate was computed for the various receiver positions assuming a target BER value, equal to 10^{-3}. The results are depicted in Fig. 6a), b) and c). It is shown that the SISO scheme can reach data rates as high as 0.6Gb/s while in less favorable positions the achieved bit rate is between 0.2 and 0.4Gb/s. The use of STBC schemes increases the maximum bit rate to 1.5Gb/s and 3Gb/s for the 2×1 and 2×2 schemes

respectively. The above results demonstrate the usefulness of OSTBC in gigabit coherent IR wireless systems in terms of the achievable bit rate.

D. PWM dimming for DMT-based visible light communication systems

The proposed VLC system is illustrated in Fig. 7 [28].

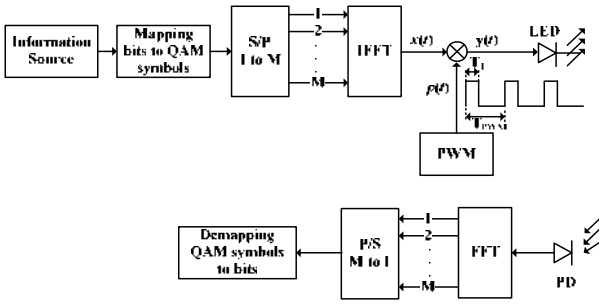


Fig. 7. Basic blocks of a PWM-DMT VLC system.

The data bits to be transmitted are converted into a sequence of symbols, using a QAM constellation mapper. The produced symbols are fed to a serial to parallel (S/P) converter and undergo an inverse Fourier transform (IFFT) operation to form the multicarrier signal $x(t)$. The generated multicarrier signal is multiplied by a periodic PWM pulse train $p(t)$ with a duty cycle of $d = T_1/T_{\text{PWM}}$, where T_1 is the duration of the PWM pulse and T_{PWM} is the period of the PWM signal. The dimming level δ is determined by $\delta = 1\text{-}d$. The resulting composite signal $y(t)=x(t)p(t)$ is the driving current of the LED. The transmitted signal $y(t)$ impinges on the photodiode receiver, undergoes an FFT operation and the resulting baseband signal is fed to a parallel to serial converter. The generated symbols are estimated using appropriate detection schemes.

The PWM signal is periodic and can therefore be expressed in terms of a Fourier series,

$$p(t) = \sum_{n=-\infty}^{\infty} C_n e^{j2\pi nt/T_{PWM}} \tag{16}$$

where C_n are the Fourier coefficients of $p(t)$. If $p(t)$ is rectangular, then C_n will be exhibiting a sinc-like dependence with respect to n. Therefore, one can show that the spectrum of the signal waveform $y(t)$, which is in general given by:

$$Y(f) = \int_{-\infty}^{+\infty} dt y(t) e^{-j2\pi ft} \tag{17}$$

can be expressed as:

$$Y(f) = \sum_{n=-\infty}^{+\infty} C_n X(f - nf_{\text{PWM}}) \tag{18}$$

According to (18), and much as in the case of digital signal sampling, the spectrum of $y(t)$ is comprised of a weighted sum of displaced versions of the original spectrum $X(f)$. In the case of a DMT signal, $X(f)$ can be assumed to be contained inside $f\in[-B\ B]$, where B is approximately the frequency of the highest subcarrier $f_{M\text{-}1}$. Consequently, if the PWM rate f_{PWM} is chosen to be twice the bandwidth B of the DMT signal ($f_{PWM} \geq 2\text{B} \cong 2f_{M\text{-}1}$) then the spectral components $C_nX(f\text{-}nf_{PWM})$ will experience no aliasing. In the case of the PWM waveform considered in Fig. 8a), $C_0 = T_1/T_{\text{PWM}} = d$, and consequently, when $f_{PWM} \geq 2\text{B} \cong 2f_{M\text{-}1}$,

$$Y(f) = X(f)d \tag{19}$$

for $f\in[-B\ B]$. According to (19), the spectrum remains unchanged in shape upon dimming, as suggested by Fig. 8b), and hence, if $F = f_{\text{PWM}}/f_{M\text{-}1} \geq 2$, one expects no interference between the subcarriers. This is an important conclusion and will be used to justify the enhanced system performance obtained for $F \geq 2$.

In order to gain a better insight on the system's performance the BER was calculated. For this analysis we resorted to Monte Carlo simulations. The system performance was investigated in terms of the achieved BER for various dimming levels. Again, $M = 32$ and

N = 16 are considered, with the last subcarrier positioned at 500 kHz. The employed detection scheme takes into account the self-interference term and should produce no errors in the case of infinite signal to interference ratio. However, it requires knowledge of the timing as well as the dimming level at the receiver. Taking into account the the ambient light noise the mean value of the BER was computed and it is shown in Fig. 9 for a synchronized system. An SNR of 45 dB, corresponding to a transmit optical power of 0.6W and a noise power spectral density of ~10^{-11} A^2/Hz, higher than in [29] was assumed, and two dimming levels (80% and 20%) were considered. The mean BER reaches low values in the region where the relative rate F is approximately larger than 2.

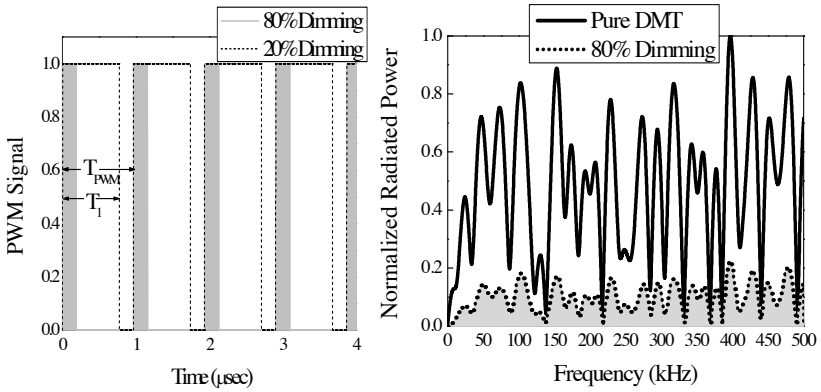


Fig. 8. a) PWM signal for two different dimming levels, i.e. 80% and 20%. b) Spectrum of composite PWM-DMT signal and pure DMT signal for F=2.

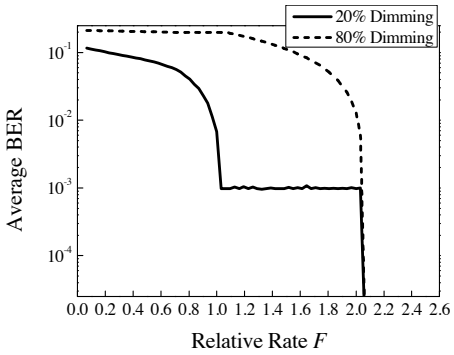


Fig. 9. Mean BER for synchronized signal, detection scheme B for various dimming levels with noise.

There is also another important implication regarding the illumination performance of the system. The optical power emitted by the LED is proportional to $y(t)$, and (19) suggests that in the frequency regime $|f_L| < 200$ Hz, where the human eye may sense optical power variations [30], incorporation of PWM will not distort the spectrum of the original signal. Consequently, PWM will not introduce any significant additional flickering.

To further investigate whether LED light flickering is induced in a PWM/DMT VLC system, we consider the low frequency component $y_{LOW}(t)$ corresponding to the part of the spectrum $Y(f)$ residing inside $f \in [-f_L f_L]$,

$$y_{\mathrm{LOW}}(t) = \int_{-f_L}^{f_L} Y(f) e^{-j2\pi ft} dt \tag{20}$$

The flickering factor C_F, defined as:

$$C_{\mathrm{F}} = \frac{\max\{|\, y_{\mathrm{LOW}}(t)\,|\} - \min\{|\, y_{\mathrm{LOW}}(t)\,|\}}{Y(0)} \tag{21}$$

where $Y(0)$ is the DC component of the signal $y(t)$, is a measure of the LED light flickering.

The values of the average C_F were calculated numerically based on Monte Carlo simulation. 100 iterations were considered, and on each, 31 QAM symbols s_m were randomly chosen, one for each active subcarrier. The DMT waveform $x(t)$ was then calculated and was multiplied by the PWM signal $p(t)$ to estimate $y(t)$. The spectrum $Y(f)$ of

$y(t)$ was calculated using a discrete Fourier transform (DFT) in the low frequency regime, $f \in [-f_L\ f_L]$. Performing an inverse DFT according to (20), we then calculate $y_{LOW}(t)$ and estimate the average flickering factor C_F obtained from each of these 100 waveforms. In Fig. 10 the relative deviation of flickering factors with respect to the undimmed case, are depicted, for δ = 80% and δ = 20%. This ratio appears to be very low, in the order of $\cong 10^{-6}$, and hence no perceivable LED light flicker is expected.

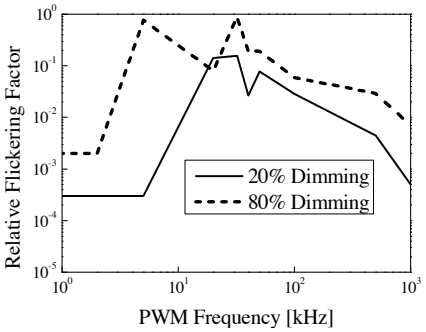


Fig. 10. Relative flickering factor for different PWM line rates when 20 and 80% dimming is considered.

III. Conclusions

In this work, the potential of indoor optical wireless systems was examined for data rates higher than 100 Mb/s. It was shown that diffuse infrared systems can reliably carry traffic of 100Mb/s inside a medium size room when appropriate equalization techniques are employed. In addition MIMO optical wireless schemes were investigated for intensity modulated systems based on STBC and repetition coding. It was shown that STBC does not provide significant performance gains while in contrast, repetition coding based MIMO could be used to relax the power constraints in future optical wireless LANs. On the other hand, STBC techniques can be used to increase the capacity of coherent optical wireless systems, improve their coverage and decrease the required optical power at the transmitter. In an effort to further enhance the performance of indoor IR systems the use of more than two transmit elements (N_T>2) would also constitute an interesting and viable solution. Such arrangements could further alleviate the high power requirements of IR systems. An important implementation issue to consider in these systems is the estimation of the channel coefficients required in symbol demodulation. Another issue to consider is the incorporation detection schemes that can allow arbitrary signal/local oscillator polarizations. These issues should be the subject of future investigation.

The implications of PWM dimming in the performance of a DMT-based VLC system were also presented. The influence of PWM in the spectrum of the DMT waveform was analyzed both theoretically and numerically. Closed-form formulas were derived enabling the estimation of the signal-to-interference ratio at the receiver. Numerical Monte Carlo simulations were used to estimate the BER of the system. The results showed that reliable communication is only possible when the PWM samples the DMT waveform at a rate faster than twice the highest subcarrier frequency of DMT. For slower PWM rates, there is significant spectral aliasing leading to prohibitively large subcarrier interference. The analysis was extended to account for ambient light noise and the main conclusions remained unaltered. Since the PWM rate should remain well within the 3dB LED bandwidth, the need to use a PWM signal twice as fast as the frequency of the last subcarrier ultimately sets a limit on the amount of information that can be transmitted over the system. The results also showed while there is no inherent need to synchronize the PWM and the DMT waveforms, synchronizing DMT to PWM is making the BER performance much less independent of clock jitter and the like. It the case of “oversampled” DMT we also showed that combining DMT and PWM does not increase the inherent flicker of PWM-dimmed light.

References

[1] F.R. Gfeller and U. Bapst, “Wireless in-house data communication via diffuse infrared radiation,” *Proc. IEEE*, vol. 67, pp. 529–551, November 1979.

[2] A.C. Boucouvals, “Indoor ambient light noise and its effect on wireless optical links,” Proc. *Optoelectronics IEE*, vol. 143, pp. 334–338, December 1996.

[3] A. J.C. Moreira, R.T. Valadas, and A.M. de Oliveira Duarte, “Characteristization and modeling of artificial light interference in optical wireless communication systems,” *Proc. IEEE Personal Indoor and Mobile Radio Communications (PIMRC ‘95)*, vol. 1, pp. 326–331, September 1995.

[4] A. J.C. Moreira, R.T. Valadas, and A.M. de Oliveira Duarte, “Optical interference produced by artificial light,” *Wireless Networks*., vol. 3, pp. 131–140, May 1997.

[5] G.W. Marsh and J.M. Kahn, “50-Mb/s diffuse infrared free-space link using on-off keying with decision feedback,” *IEEE Photonics Technology Letters*, vol. 6, pp. 1268–1270, October 1994.

[6] D. C. O’Brien, L. Zeng, H. Le-Minh, G. Faulkner, O. Bouchet, S. Randel, J. Walewski, J. A. R. Borges, K.-Di. Langer, J. Grubor, K. Lee and E. T. Won, “Visible Light Communications”, White-Paper in Wireless World Research Forum 20 (WWRF 20), Ottawa, Canada, Apr. 2007

[7] H.Le-Minh, D. C. O’Brien, L. Zeng, O. Bouchet, S. Randel, J. Walewski, J. A. R. Borges, K.-Di Langer, J. Grubor, K. Lee and E. T. Won, “Short-range Visible Light Communications”, Proceedings of Wireless World Research Forum 19 (WWRF 19), Chennai, India, Nov. 2007.

[8] T.Komine, Y. Tanaka, S. Haruyama and M. Nakagawa, “Basic Study on Visible-Light Communication using Light Emitting Diode Illumination,” Proc. of the 11th Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2000), London, US, pp. 1325-1329, 2000.

[9] J. Grubor, O. C. Gaete Jamett, J. W. Walewski, S. Randel, K. –D. Langer, “High-Speed Wireless Indoor Communication via Visible Light,” ITG Fachbericht, 2007, pp.203-208.

[10] [12] S. Randel, F. Breyer, S. C. J. Lee and J. W. Walewski, “Advanced modulation schemes for short-range optical communications,” Journal of Selected Topics in Quantum Electronics, September/October 2010.

[11] [13] J. A. C. Bingham, “Multicarrier modulation for data transmission: An idea whose time has come,” IEEE Communications Magazine, vol. 28, pp. 5-14, May 1990.

[12] [14] J. Vucic, C. Kottke, S. Nerreter, A. Buttner, K.-D. Langer, J. W. Walewski, “White Light Wireless Transmission at 200+ Mb/s Net Data Rate by Use of Discrete-Multitone Modulation,” IEEE Photonics Technology Letters, Vol. 21, Issue 20, pp. 1511 – 1513, Oct.15, 2009.

[13] [15] Y. Zhang, Z. Zhang, Z. Huang, H. Cai, L. Xia and J. Zhao, “Apparent brightness of LEDs under different dimming methods,” Proc. SPIE, Vol. 6841, 2007.

[14] [16] S. Muthu, F. J. Schuurmans, M. D. Pashley, “Red, Green, and Blue LED based white light generation: Issues and control,”37th Annual IEEE-IAS meeting, Vol. 2, 327-333,2002.

[15] F.J. Lopez-Hernandez, R. Perez-Jimenez, and A. Santamaria, “Modified Monte Carlo scheme for high-efficiency simulation of the impulse response on diffuse IR wireless indoor channels,” *Electronics Letters*, vol. 34, pp. 1819–1820, September 1998.

[16] J.B. Carruthers and J.M. kahn, “Angle diversity for non-directed wireless infrared communication,” IEEE Trans. On Comm., vol. 48, pp. 960–969, June 2000.

[17] J.G. Proakis, *Digital Communication*, 4th ed., New York: Mc Graw-Hill, 1995, pp. 601-626.

[18] G. Ntogari, T. Kamalakis and T. Sphicopoulos, “Performance Analysis of Non-Directed Equalized Indoor Optical Wireless Systems”, CSNDSP08 Graz, Austria, 2008.

[19] G. Ntogari, T. Kamalakis and T. Sphicopoulos, 'Performance Analysis of Decision Feedback and Linear Equalization schemes for Non-Directed Indoor Optical Wireless Systems', Journal of Communications, Vol 4, No 8,pp. 565-571, 2009.

[20] S. M. Alamouti, “A Simple Transmit Diversity Technique for Wireless Communications,” IEEE J. Select. Areas Comm,, vol. 16, pp. 1451-1458, October 1998.

[21] P. Djahani and J. M. Kahn, “Analysis of Infrared Wireless Links Employing Multibeam Transmitters and Imaging Diversity Receivers”, *IEEE Trans. on Comm*., vol. 48, no.12, pp.2077-2088, December 2000.

[22] J. Grubor, S. Ch. J. Lee, K-D. Langer, T. Koonen, and J. W. Walewski, “Wireless High-Speed Data Transmission with Phosphorescent White-Light LEDs,” Proceedings 33rd European Conference and Exhibition on Optical Communication, Vol. 6, Post-Deadline Papers, PD3.6, 2007.

[23] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press 2003.

[24] G. Ntogari, T. Kamalakis, and T. Sphicopoulos, "Performance analysis of space time block coding techniques for indoor optical wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 1545-1552, 2009.

[25] M. Jafar, D. C. O’Brien, C. J. Stevens, and D. J. Edwards, “Evaluation of coverage area for a wide line-of-sight indoor optical free-space communication system employing coherent detection,” *IET Commun*. vol. 2, pp. 18–26, 2008.

[26] D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, “Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation,” *Journal of Lightw*ave *Technol*., vol. 24, no. 1, January 2006.

[27] G. Ntogari, T. Kamalakis, and T. Sphicopoulos, “Analysis of indoor multiple input multiple output coherent optical wireless systems,” accepted for publication at IEEE *J. of Lightw. Technol*., 2011.

[28] G. Ntogari, T. Kamalakis, J. Walewski, T. Sphicopoulos, 'Combining Illumination Dimming Based on Pulse-Width Modulation With Visible-Light Communications Based on Discrete Multitone', IEEE/OSA Journal of Optical Communications and Networking, 3 (1), pp. 56-65, 2011.

[29] J. Grubor, S.Randel, K.-D. Langer, and J. W. Walewski, “Broadband Information Broadcasting Using LED-Based Interior Lighting,” Journal of Lightwave Technology, vol. 26, no.24, December, 2008.

[30] IEEE P802.15.7 Standard, “Part 15.7: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Visible Light Wireless Personal Area Networks (WPANs),” The Institute of Electrical and Electronics Engineers, Inc. 2009.

High Level Multimodal Fusion and Semantics Extraction

Thanassis Perperis**

National and Kapodistrian University of Athens
Department of Informatics and Telecomunications
a.perperis@di.uoa.gr

Abstract. This thesis on multimodal fusion and semantics extraction, focuses on automated detection and annotation of harmful content in video data. The aim is not only to reason out the existence of violence or not (i.e. the binary problem), but also to determine the type of violence (e.g. fight, explosion, murder). Acknowledging the lack of knowledge representation and reasoning approaches for the problem at hand, we propose a semantic fusion approach that combines low to mid level modality specific semantics through ontological and rule reasoning. A major part of the proposed framework is the movie segmentation into meaningful and easy to handle units. We evaluate a set of shot boundary detection approaches combined through a majority voting scheme. In the sequel, state of the art classification methods are employed to extract audio and visual mid level semantics. The segmentation and modality specific analysis algorithms instantiate the corresponding video structure and modality specific ontologies developed in the context of the knowledge engineering framework. A set of consecutive and interleaved ontological and SWRL rule reasoning steps map sets and sequences of extracted low to mid level semantics into higher level concepts represented in the harmful content domain ontology. We present the involved ontologies, the corresponding SWRL rule sets and the reasoning mechanism in detail. Finally we present the evaluation of the proposed approach in a preanotated movie dataset, compare its results with the single modality approaches and a kNN late fusion meta classifier. We comment on the higher level semantics extraction ability and evaluate a set of extensions employed in the basic structure of the framework. The extensions concern the development of a scene detection module that combines markov clustering with SQWRL queries, the incorporation of existing rating and movie genre metadata in the violence identification procedure and the detection of pornography.

Keywords: Harmful Content Detection, Semantic Video Analysis, Knowledge Engineering, Ontologies, Rules, Reasoning

** Dissertation Advisor: Professor Sergios Theodoridis

1 Introduction

Detecting Harmful Content in video data, which are complex in nature, multimodal and of significant size is not an easy task and requires extensive and efficient analysis. Although quite different, in terms of the exploited modalities and methodology specific details, most of the proposed approaches in the literature fall in the *pattern recognition* discipline, following either or single- or a multi modal- approach. In the multimodal approach, the modality fusion is performed either at the feature level (i.e. early fusion) or at the decision level (i.e. late fusion).Examining the reported research on harmful content detection, we conclude that the extracted semantics are not at the desired level for higher level harmful content filtering applications. Most approaches tackle either the binary or constrained instances of the multiclass problem and are rather tailor made. In addition, there is a lack of interest on incorporating automatic annotation processes in terms either of MPEG-7[1] or of ontological representations. In this thesis we tackle violence and pornography detection by means of a unified approach exporting various levels of semantic abstractions, employing state of the art audio, visual and textual mid-level concept detectors, using domain knowledge, ontologies and reasoning for higher level semantics inferencing, annotation and filtering. Deploying ontologies and semantic web trends in the context of video analysis for automatic annotation and filtering is a very challenging task. Current ontological approaches appear mostly to enhance semantic searching applications in simplified and constrained domains like medical [1, 2], sports [3, 4] and surveillance [5] videos whereas limited work appears in the domain of movies and TV series for harmful content detection. An ad hoc ontological approach, exploiting Video Event Representation Language (VERL) [6] along with Video Event Markup Language (VEML), for surveillance, physical security and meeting video event detection applications, emerged as a result of the ARDA event taxonomy challenge project [7]. In addition, an interesting approach aiming towards extracting and capturing the hierarchical nature of actions and interactions, in terms of *Context Free Grammars (CFG)* rather than ontologies, from raw image sequences appears in [8]. To the best of our knowledge, the only approaches, directly employing multimedia and domain ontologies for multimodal video analysis, in the context of harmful content identification, are our works for violence detection presented in [9, 10] and [11, 12].

2 Knowledge-based framework for violence identification in movies

The presented ontological (Fig. 1) approach[2] is a direct application of the proposed framework in [9], further elaborating, extending and implementing

[1] http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

[2] This work has been supported by the Greek Secretariat for Research and Technology, in the framework of the PENED program, grant number TP698.

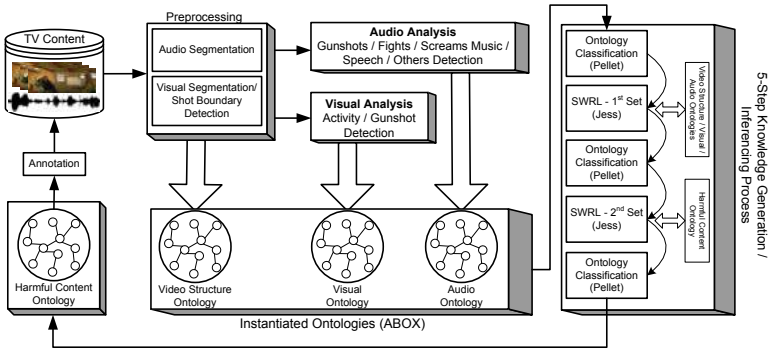


Fig. 1. Proposed Methodology For Violence Detection

each of the involved modules. Our aim is not to devise high quality low level analysis processes, but to propose an open and extendable framework, combining existing single modality low to mid-level concept detectors with ontologies and SWRL[3] rules, to identify and formally annotate an *extensive range of complex violent actions* in video data. We identify as major processes of the system, a preprocessing/segmentation step, a visual analysis step, an audio analysis step, each one interconnected with the corresponding data ontology, an inferencing procedure and a harmful content domain ontology. Speaking in terms of knowledge engineering, low level analysis extracts *basic facts - basic truth* that holds for the corresponding TV/video content data, while ontologies define *complex* and *terminological* facts that hold for the examined domain in general. Thus an explicit knowledge base is formed and the scope of the 5-step inferencing procedure is to draw new implicit conclusions/knowledge during each step. In the following, we further elaborate on the implementation decisions adopted for each of the aforementioned processes.

2.1 Preprocessing - Segmentation Semantics

Preprocessing tackles the task of temporal audio visual segmentation and feeds the corresponding low level analysis algorithms. The role of segmentation is three-fold: 1) Define the temporal annotation units, 2) Feed low level analysis with temporal segments of predefined duration, 3) Preserve a common time reference and extract sequence and overlapping relations, among visual and audio segments, for the temporal reasoning procedures. Thus, both modalities are initially segmented into fixed duration segments (1-sec as the minimum event duration), as defined from the low level modality specific algorithms, and then grouped into the corresponding shots to preserve the synchronicity and time reference, further forming the annotation units. The task of shot

[3] Employed Ontologies and SWRL rules are available through http://hermes.ait.gr/PenedHCD/

boundary detection is performed by means of a majority voting local content adaptive thresholding approach. In short, the grayscale pixel difference, the RGB pixel difference, the RGB and HSV histogram differences, the RGB and HSV histogram similarities based on a combined measure, the edge change ratio and the average motion through motion vectors are computed for each pair of consecutive frames. Further the mean values of the afforementioned measures are computed on a 5-D temporal window. If the current difference is maximum in the examined window and is greater than twice the mean window value, a shot cut is detected for the examined measure. Finally a shot cut is detected when the majority of approaches satisfies the adaptive thresholding condition. Experimentation of the shot boundary detection module on a violent movie dataset achieves 95.18% for precision and 91.01% for recall.

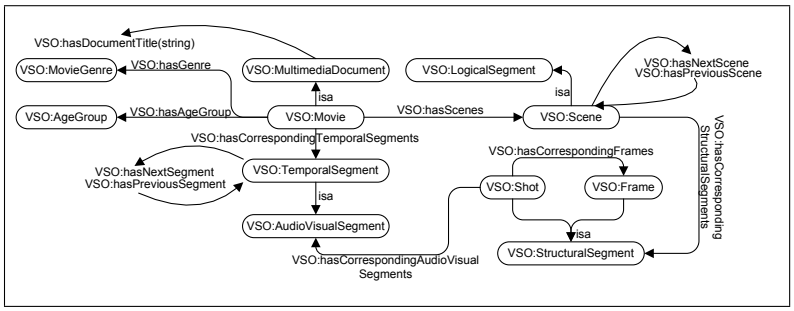


Fig. 2. Video Structure Ontology

For interoperability of the segmentation process reasons with modality specific and domain ontologies, the overall methodology demands for a Video Structure Ontology (Fig. 2), capturing temporal and structural semantics along with authoring information. In short, an example of the core classes is *VSO:MultimediaDocument*[4] that defines the class of documents containing combinations of audio, visual and textual information. Although our final target is to provide complete ontological representation for a multitude of multimedia document (e.g. streaming videos, web pages, images, audio recordings) currently, the only implemented subclass is the one representing movie/TV series content, defined to include individuals instantiating the corresponding property axioms with at least one temporal and at least one structural segment. In addition, Datatype properties (i.e. VSO:hasDocumentTitle, VSO:hasTotalFrameNumber, VSO:hasFrameRate) capture authoring information. For further interoperation with existing metadata annotations (e.g. TV-Anytime) providing overall content description, in terms of intended age groups or genre classification, corresponding axioms *VSO:hasAgeGroup*, *VSO:hasGenre* were included in the ontological

[4] VSO: Stands for *Video Structure Ontology* and is used as prefix for every element of the Structure Ontology

definition. *VSO:TemporalSegment* categorizes segments (in our case of 1-sec duration) exploiting either the auditory or visual modality or both to convey meaning associated with a temporal duration. Structurally every video is a sequence of shots (i.e. single camera capturing) and every shot is a sequence of frames. Logically, every video is a sequence of semantically meaningful scenes, each one composed of consecutive shots. Thus *VSO:StructuralSegment* defines content's elementary structural and *VSO:LogicalSegment* subsumed logical segments. The last two classes and their property axioms serve as the main interconnection point with the low level analysis algorithms and are the first to be instantiated, initiating thus the inferencing procedure. Obviously there is a strong interconnection of the video structure ontology with an MPEG-7 one. Thus we have incorporated, with property axioms, an extended with the MPEG-7 audio part version of Jane Hunter's ontology [13], aiming further for automated MPEG-7 based annotation.

2.2 Audio Visual Semantics

To optimally combine multimedia descriptions with the violence domain ontology, the knowledge representation process has further defined modality violence ontologies (audio, visual) that essentially map low-level analysis results to simple violent events and objects (medium-level semantics). Adopting a top down approach, the modality specific ontologies comprise an important "guide" for low level analysis algorithms. Namely, they define what to search or try to extract from raw data. Although concept detectors using statistical, probabilistic and machine learning approaches, are extensively studied in the literature, it is not yet possible to extract whatever a knowledge engineering expert prescribes. Consequently, the corresponding ontologies contain a broader set of potential for extraction concepts, further prioritizing low level analysis research towards developing novel concept detectors. Taking under consideration the intrinsic difficulties of low level analysis, providing erroneous results in some cases, the long-term target is to connect a broad set of concepts/events/object detectors with the modality specific ontologies, providing thus a cooperative mechanism towards increasing the detection accuracy.

Visual Semantics The Visual Ontology (Fig. 3) defines, in a hierarchical way, the set of objects/events - possibly related with harmful content - identified during visual analysis. Although a much broader set of low to mid level semantics is already defined, for interoperability with the employed visual analysis algorithms reasons, we focus our attention on the actually extracted visual clues. In particular, each fixed duration segment is classified in one of three activity (low, normal, high) and two gunshot (gunshot, no-gunshot) classes. Thus, the simple human events and visual objects ontological classes of

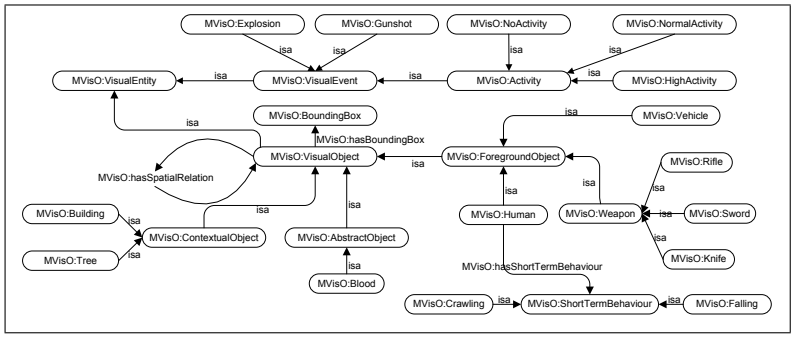


Fig. 3. Visual Ontology - Objects And Events

interest are *MVisO:HighActivity*[5], *MVisO:NormalActivity*, *MVisO:NoActivity*, *MVisO:Gunshot*, *MVisO:noGunshot* and *MVisO:Face*.

The activity detection process initially computes, in a per segment basis, the *average overall motion*, *variance of motion* and *average degree of overlap of the detected people (i.e. faces)* features. The former ones are computed by the corresponding *motion vectors* while the latter by the face detection and tracking component. The face detection component employs a set of Haar-like features and a boosted classifier to detect people in the scene while the tracking component employs the hierarchical fusion algorithm devised in [14]. To increase face detection robustness, a histogram based skin detection algorithm filters-out objects with minimum skin content. Finally, the classification process is performed by means of a weighted kNN (k-Nearest Neighbor) classifier that computes the segment likelihood to attain high, normal or no activity content. Measuring abrupt changes in the illumination intensity can provide valuable hints on the existence of fire, explosions or gunshots. Thus, the *maximum luminance difference* and *maximum luminance interval* features are used to train a distinct weighted kNN classifier discriminating between gunshot and no-gunshot segments. For further details on the visual analysis components the reader is referred to [11].

Audio Semantics Additional clues increasing the harmful content detection accuracy exist in the auditory modality. Contrary to visual, in audio semantics the ontological definition covers the full set of extracted mid level semantics. Thus the audio classes (Fig. 4) of interest are *MSO:Screams*[6], *MSO:Speech*, *MSO:Gunshot*, *MSO:Fights*, *MSO:SharpEnviromentalSound* and *MSO:SmoothEnviromentalSound* . The audio analysis process involves a variant of the "One-vs-All" (OVA) classification scheme presented in [15], on a segment

[5] MVisO: Stands for *Movie Visual Ontology* and is used as prefix for every element of the Visual Ontology

[6] MSO: Stands for *Movie Sound Ontology* and is used as prefix for every element of the Sound Ontology

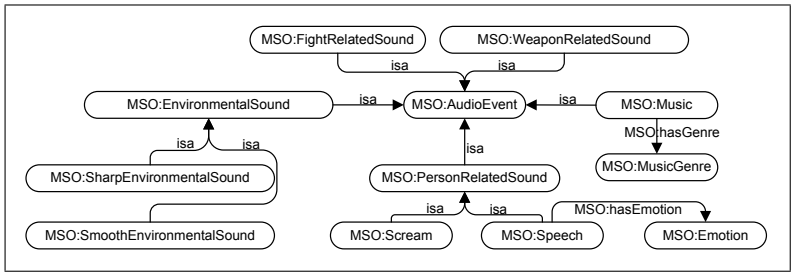


Fig. 4. Audio Ontology

basis, in order to generate a sequence of three violent (*Gunshots*, *Fights* and *Screams*) and four non - violent (*Music*, *Speech*, *Others1*: environmental sounds of low energy and almost stable signal level like silence, background noise, etc., and *Others2*: environmental sounds with abrupt signal changes like thunders or closing doors) audio class probabilities for each segment.

2.3 Domain Ontology Definition

An effective formal representation of the harmful content domain knowledge in all its complexity, abstractness and hierarchy depth, to drive corresponding acts detection, has never been attempted before. We have made a step forward towards this direction. The ontology definition has resulted from an extended investigation through close observation of mostly violent and some pornographic acts in video data collected from TV programs, movies, streaming videos, news and security camera captures. The Harmful Content Domain Ontology[7] (Fig. 5) includes the definition of numerous high level concepts as a set of related spatiotemporal entities (i.e. actions, events, objects) irrespective of low level analysis algorithms. Taking under consideration the inability of employed audio and visual analysis algorithms to extract the broad set of violence related modality specific semantics (i.e. guns, swords, vehicles, contextual objects, body parts, emotional state, simple actions, events, etc.) we are forced to limit our description and experimentation on the set of attained mid-level semantics. In addition due to *open world reasoning* in OWL, we cannot identify non-violence directly using simple *negation as failure* reasoning. Therefore the domain ontology further aims on the identification of a number of non harmful classes like dialogue, actions/activity and scenery.

2.4 Inferencing Procedure

Having the extracted low- to mid- level semantics, and the corresponding loosely coupled, using common terms, equivalent classes and object property

[7] HCO: Stands for *Harmful Content Ontology* and used as prefix for every element of the Domain Ontology.

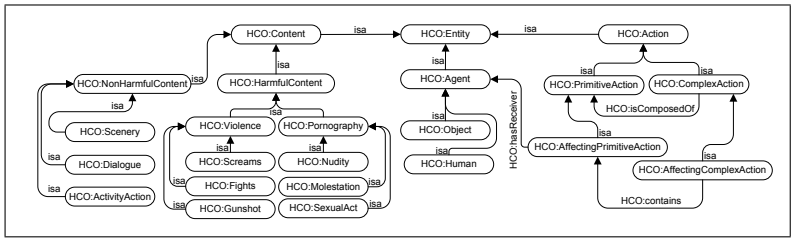


Fig. 5. Harmful Content Ontology

axioms, ontological descriptions, we have to tackle with the issue of *fusing and interchanging semantics from different modalities*, towards inferring more complex, abstract and extensive violent cases represented in the domain ontology. Thus there is a need for a cross-modality ontological association mechanism between modality specific and domain ontologies, further increasing the semantic extraction capabilities of low level analysis. Towards this direction we investigate the usage of SWRL rules[8] combined in a 5-step inferencing procedure with ontological reasoning. SWRL (a Horn-like rule language combining OWL with RuleML), on the one hand, can reason about OWL instances (individuals) in terms of OWL classes and properties and is mostly used to increase the OWL expressivity. Ontological reasoning, on the other hand, implements consistency checking, classification and instance checking services, which is usually achieved using one of the existing reasoners (in our case Pellet[9]). Although manual SWRL rule construction is a tough procedure, especially for such a complicated domain like harmful content detection, we explore the potential for cross-modality representation of violent spatio-temporal relations and behavioral patterns. Since SWRL and OWL do not yet formally support reasoning under uncertainty, we attempt to capture such semantics by simply thresholding the corresponding datatype relations with SWRL built-in axioms.

Speaking in terms of Description Logic (DL), the involved ontologies define the conceptual representation of the domain in question, forming thus the *Terminological Box (TBox)* of our knowledge base. Taking into consideration that the inferencing procedure attempts to reclassify modality individuals to the violence domain ontology, the definition of the *Assertional Box (ABox)* - the set of individual definitions in terms of TBox's concepts, roles and axioms - is mandatory. In the examined field of application, the ABox derives directly from the segmentation and audio/visual analysis algorithms, forming the basic facts in terms of individuals representing segments, frames, events and objects along with the corresponding low level numerical values and occurrence/accuracy probability results. Essentially, the instantiated model captures existing and extracted knowledge for the TV content in question. Based on this explicit knowledge, the 5-step inferencing procedure is applied. Each

[8] http://www.w3.org/Submission/SWRL
[9] http://clarkparsia.com/pellet/

ontological reasoning step draws implications that could possibly trigger the execution of a new rule set. Similarly, each rule execution implies fresh knowledge exploited by subsequent ontological reasoning steps. Namely, the process is as follows:

Step-1: During ontology instantiation, individuals are subclassed under *owl:Thing*, thus the main purpose of this step is to check the consistency of the instantiated VSO model and assert each individual's initial class, according to TBox's necessary and sufficient conditions.

Step-2: The first SWRL rule set, composed of 35 rules, is applied. Datatype property axioms expressing arithmetic probabilities and numerical values are translated to object property axioms, instantiated with classes and instances from the modality specific ontologies (e.g. assign the winner class). In addition, instantiation of the extracted video structure is performed (i.e. shots are related with the corresponding segments and segments with the corresponding frames).

Step-3: Consistency checking and classification services on the implied modality specific models are applied.

Step-4: The second SWRL rule set composed of 39 rules is applied. Implied audio and visual mid level semantics are combined using common sense logic, spatiotemporal relations and simplified conclusions drawn from low level analysis modules (i.e. confusion matrices) for cross-modality reasoning, further mapping video segments in one of the domain ontology classes.

Step-5: In this final step, consistency checking and classification services on HCO are applied to infer violent and non violent segments (instances reclassify from children to parents) on the one hand and extended semantics (instances reclassify from parents to children) on the other hand. Since the first classification case is straightforward (e.g. every fight segment is also a violent one) we will further describe the second case using a simple example.

2.5 Implementation and Experimentation

The presented ontological approach was developed using Matlab and the OpenCV library[10] for audio/visual feature extraction and classification, Protégé[11] for the definition of ontologies and SWRL rules, Pellet and Jess[12] for ontology reasoning services and rules execution and finally Jena semantic web framework[13] for ontologies instantiation and synchronization of the knowledge generation lifecycle. For evaluation purposes, 50 videos have been extracted from 10 different movies. The total duration of the test data is 2.5 hours. The video streams have been manually annotated by three persons, using the Anvil[14] video annotation research tool to generate ground truth data for performance evaluation. According to manual annotations 19.4% of the data was of violent content.

[10] http://sourceforge.net/projects/opencvlibrary/
[11] http://protege.stanford.edu
[12] http://www.jessrules.com/
[13] http://jena.sourceforge.net/
[14] http://www.anvil-software.de/

	Recall	Precision	F_1	Mean Accuracy
Audio-binary	82.9%	38.9%	53%	61%
Visual-binary	75.6%	34%	46.9%	54.8%
Violence Inference	91.2%	34.2%	50%	62.7%
Fights Inference	61.6%	68.2%	64.8%	64.9%
Screams Inference	41.4%	33.5%	37.1%	37.4%
Shots-Explosions Inference	63.3%	38.2%	47.6%	50.7%

Table 1. Segment based Binary and Multiclass Detection Performance Measures

Tables 1 and 2 demonstrate the achieved experimental results for the binary and multiclass violence detection problem in a per segment (1-sec duration) and in a per shot basis. For comparison with the single modality approaches purposes in table 1, we further demonstrate the achieved performance of low level audio and visual analysis algorithms for the binary problem. The performance measures are computed by means of ***P****recision* (i.e. the number of correctly detected violence segments, divided by the total number of detected violence segments), ***R****ecall* (i.e. the number of correctly detected violence segments divided by the total number of **true** violence segments), *Mean Accuracy* and F_1 *measure* ($F_1 = \frac{2 \cdot P \cdot R}{P+R}$).

	Recall	Precision	F_1	Mean Accuracy
Violence Inference	61.07%	68.80%	64.7%	64.93%
Fights Inference	68.72%	89.9%	77.89%	79.31%
Screams Inference	25.0%	41.17%	31.10%	33.08%
Shots-Explosions Inference	89.39%	40.68%	55.91%	65.03%

Table 2. Shot based Binary and Multiclass Detection Performance Measures.

3 Enhanced Knowledge-based Framework

In this section we briefly sketch a set of elementary extensions deployed in the initial ontological framework. In order to increase the semantics detection ability from the one hand and the detection accuracy from the other hand we feed the semantic framework with the results of three pairs of Support Vector Machines and Multilayer Perceptron audio classifiers, trained using Principal Components Analysis on 160-D feature vectors extracted from the initial 7 class dataset, a 9 class music genre dataset and a 4 class pornography related dataset. Further epxloiting the shot boundary information along with log average luminance and a trained gaussian skin classifier for explosions/fire/gunshots components and large skin areas for pornography detection. Importing those mid level demands for new SWRL rule construction and results in increased performance for the violece detection case. Experimentation on a pornographic movie dataset results in 87.01% and 92.44% recall rate for the 1-sec and the shot case respectively. Finally we employ unsupervised markov clustering along with SQWRL queries to tackle scene detection. We define scenes as consecutive shots with audiovisual

and semantic coherence. Audio visual coherence is achieved through markov unsupervised clustering applied on a complete weighted shot graph. Vertex distance is defined as $w_{i,j} = dist(\bar{x_i}, \bar{x_j}) = (1.0 - e^{\gamma \cdot \|\bar{x_i} - \bar{x_j}\|^2}) \cdot e^{-\frac{(t_i - t_j)}{k}}$ where $\bar{x}$ is a 38-D multimodal feature vector. The MCL algorithm retrieves graph node clusters of maximum similarity and instatiates the coresponding video structure ontology class. At the end of the knowledge generation lifecycle a set of SQWRL queries is applied to retrieve the set of consecutive semantically equivalent shot instances that belong in the same cluster and define a scene.

4 Conclusions

In this thesis we have proposed a complete ontological framework for harmful content identification and experimentally evaluated the violence detection case. The presented approach performs better than the employed audio and visual ones, for the 1-sec segment based binary violence detection problem, by means of recall rate, and achieves a small boosting in terms of mean accuracy. The performance is significantly increased both for the binary and the multiclass problem in the shot based approach. We notice that for the *fight* class we achieve the best results whereas for the *screams* class the worst. This happens on the one hand because audio analysis algorithms produce the most accurate hints for fights identification and on the other hand because visual analysis does not actually aid screams identification. Summarizing the low value of attained results is greatly affected by the following facts: *i*) Extracted visual analysis clues are not at the desired level, since activity classification is rather generic for specific concepts identification and detection of violence related objects (i.e. guns, swords, knifes) and human actions (i.e. punch, kick) remains unfeasible. *ii*) Extracted audio and visual mid level clues are biased towards non-violence (i.e. five violent classes: audio-gunshot, screams, fights, high activity, visual-gunshot and seven non-violent: music, speech, smooth environmental sound, sharp environmental sound, no activity, normal activity, no gunshot). *iii*) Uncertain single modality results are treated as certain. We conclude that the attained results are really promising both for the binary and multiclass violence detection problem and that the main advantage of using such an ontological approach still remains the higher level semantics extraction ability, using an unsupervised procedure and common sense reasoning.

References

1. Jie Bao, Yu Cao, Wallapak Tavanapong, and Vasant Honavar. Integration of Domain-Specific and Domain-Independent Ontologies for Colonoscopy Video Database Annotation. In *Proceedings of the International Conference on Information and Knowledge Engineering*, pages 82–90, Las Vegas, Nevada, USA, 21-24 Jun. 2004. CSREA Press.
2. Jianping Fan, Hangzai Luo, Yuli Gao, and Ramesh Jain. Incorporating Concept Ontology for Hierarchical Video Classification, Annotation, and Visualization. *IEEE Transactions on Multimedia*, 9(5):939–957, 2007.

3. Marco Bertini, Alberto Del Bimbo, and Carlo Torniai. Automatic Video Annotation Using Ontologies Extended with Visual Information. In *Proceedings of the 13th ACM International Conference on Multimedia*, ACM Multimedia, Singapore, 6-11 Nov. 2005. ACM.
4. Liang Bai, Songyang Lao, Gareth J. F. Jones, and Alan F. Smeaton. Video semantic content analysis based on ontology. In *Proceedings of the 11th International Machine Vision and Image Processing Conference*, pages 117–124, Maynooth, Ireland, 5-7 Sept. 2007. IEEE Computer Society.
5. Lauro Snidaro, Massimo Belluz, and Gian Luca Foresti. Domain Knowledge for Surveillance Applications. In *Proceedings of the 10th International Conference on Information Fusion*, pages 1–6, 2007.
6. Alexandre R. J. Francois, Ram Nevatia, Jerry Hobbs, and Robert C. Bolles. VERL: An Ontology Framework for Representing and Annotating Video Events. *IEEE MultiMedia*, 12(4):76–86, 2005.
7. Bob Bolles and Ram Nevatia. A Hierarchical Video Event Ontology in OWL, ARDA Challenge Workshop Report, 2004.
8. Michael S. Ryoo and J. K. Aggarwal. Semantic Understanding of Continued and Recursive Human Activities. In *Proceedings of the 18th International Conference on Pattern Recognition*, pages 379–382, Hong Kong, 20-24 Aug. 2006. IEEE Computer Society.
9. Thanassis Perperis, Sofia Tsekeridou, and Sergios Theodoridis. An Ontological Approach to Semantic Video Analysis for Violence Identification. In *Proceedings of I-Media'07 and I-Semantics'07. International Conferences on New Media Technologies and Semantic Technologies (Triple-i: i-Know, i-Semantics, i-Media). Best Paper Award In Multimedia Metadata Applications (M3A) Workshop*, pages 139–146, Graz, Austria, 5-7 Sept. 2007.
10. Thanassis Perperis and Sofia Tsekeridou. A knowledge engineering approach for complex violence identification in movies. In *Artificial Intelligence and Innovations 2007: from Theory to Applications, Proceedings of the 4th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2007)*, volume 247, pages 357–364, Peania, Athens, Greece, 19-21 Sep. 2007. Springer.
11. Thanassis Perperis, Theodoros Giannakopoulos, Alexandros Makris, Dimitrios I. Kosmopoulos, Sofia Tsekeridou, Stavros J. Perantonis, and Sergios Theodoridis. Multimodal and Ontology-based Fusion Approaches of Audio and Visual Processing for Violence Detection in Movies. *Expert Systems with Applications*, 38(11):14102 – 14116, October 2011.
12. Thanassis Perperis and Sofia Tsekeridou. *TV Content Analysis*, chapter TV Content Analysis and Annotation for Parental Control. CRC Press, Taylor Francis, 2011.
13. Jane Hunter. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In *Proceedings of the First Semantic Web Working Symposium*, pages 261–281, Stanford, USA, 2001.
14. Alexandros Makris, Dimitris Kosmopoulos, Stavros S. Perantonis, and Sergios Theodoridis. Hierarchical Feature Fusion for Visual Tracking. In *IEEE International Conference on Image Processing*, volume 6, pages 289 – 292, San Antonio, Texas, USA, 16 Sept. - 19 Oct. 2007.
15. Theodoros Giannakopoulos, Aggelos Pikrakis, and Sergios Theodoridis. A Multi-Class Audio Classification Method With Respect To Violent Content In Movies, Using Bayesian Networks. In *Proceedings of IEEE International Workshop on Multimedia Signal Processing*, 2007.

Machine Learning in Natural Language Processing

Georgios P. Petasis*

Software and Knowledge Engineering Laboratory
Institute of Informatics and Telecommunications
National Centre for Scientific Research (N.C.S.R.) "Demokritos"
GR-153 10, P.O. BOX 60228, Aghia Paraskevi,
Athens, Greece
petasis@iit.demokritos.gr

Abstract. This thesis examines the use of machine learning techniques in various tasks of natural language processing, mainly for the task of information extraction from texts. The objectives are the improvement of adaptability of information extraction systems to new thematic domains (or even languages), and the improvement of their performance using as fewer resources (either linguistic or human) as possible. This thesis has examined two main axes: a) the research and assessment of existing algorithms of machine learning mainly in the stages of linguistic pre-processing (such as part of speech tagging) and named-entity recognition, and b) the creation of a new machine learning algorithm and its assessment on synthetic data, as well as in real world data for the task of relation extraction between named entities. This new algorithm belongs to the category of inductive grammar learning, and can infer context free grammars from only positive examples.

Keywords: information extraction, machine learning, grammatical inference.

1 Introduction

This doctoral thesis researches the possibility of exploiting machine learning techniques in the research area of natural language processing, aiming at the confrontation of the problems of upgrade as well as adaptation of natural language processing systems in new thematic domains or languages. The research is delimited in three important axes of information extraction systems:

- Part of speech recognition for the Greek language.
- Named entity recognition.
- Relation extraction between recognised named entities.

This thesis examines how machine learning methods and techniques can be exploited for the development of systems that support these tasks, which can be

* Dissertation Advisor: Constantine Halatsis, Professor

adapted more easily to new thematic domains and languages in contrast to the conventional systems that are rule based, manufactured often by experts. More specifically, this thesis researches techniques of machine learning along two main axes:

1. The application of existing techniques (both symbolic and statistical) in selected tasks of information extraction. These techniques are evaluated comparatively to each other in both the Greek and English languages. All existing machine learning algorithms that were examined require a vector of constant length as input. However the transformation of natural language into vectors of constant length is not always easy, without the use of arbitrary limits regarding the maximum number of words. This observation constituted the motivation for the creation of a new machine learning algorithm, which does not require vectors of constant length as input.
2. The development of a new machine learning algorithm, without the requirement for vectors of constant length as input. This new algorithm learns context free grammars from positive examples, with guidance via heuristics, such as minimum description length.

Regarding the first axis, named entity recognition systems were developed and evaluated, based on existing machine learning algorithms, such as decision trees and neural networks. The systems that were developed concern various thematic domains (management succession events, financial news, and juridical decisions) both in the Greek and English languages. These systems were evaluated in Greek texts, and they led to the recognition of the disadvantages and restrictions imposed by the examined algorithms, when applied on natural language data. From this analysis we concluded that one of the main problems when applying machine learning is the difficulty in managing data of variable length, as for example the information concerning all words of a sentence. On the contrary, a syntactic analyser can easily decide if a sentence (or part of a sentence) is described by a provided grammar. However, the manual development of grammars suitable for a specific task is a complex process, while the results frequently depend on the thematic domain and of course from the language. Consequently, if such a grammar can be automatically acquired with the use of machine learning, then the adaptation of systems that use such grammars to new thematic domains or languages can be considerably simplified.

The contribution of the developed systems is significant. The named entity recognition systems that were developed for the Greek language are among the first systems of their kind that have been reported in the bibliography. Simultaneously, the performance of the developed systems is satisfactory, and directly comparable to the performance of similar systems reported in the bibliography for the corresponding time period.

Regarding the second axis, and aiming at the confrontation of problems associated with the application of existing techniques, a new technique of machine learning has been developed. This new technique belongs to the category of inductive grammar learning. The main advantages of this method with respect to other machine learning methods are the ability to handle textual data, as

well as the possibility of using learned grammars in existing systems, replacing manually developed grammars. The main objective of this new technique is the automatic grammar creation, which can be used with the plethora of available syntactic parsers that have been presented in the bibliography, replacing existing (and probably manually constructed) grammars for various tasks in information extraction systems.

For applying inductive grammar learning, a new algorithm has been developed that learns grammars from positive examples only. This new algorithm can infer context free grammars, and it has been based on the existing algorithm GRIDS [1], improving both the used heuristic, as well as the search process in the space of possible grammars, increasing simultaneously the applicability of the new algorithm to bigger collections of data. The requirement for the algorithm to function only with positive examples emanates from the frequent absence of negative examples in the area of natural language processing. It should be noted that the presence of negative examples constitutes a necessary condition for the operation of most existing grammatical inference algorithms. The design of this new algorithm has been done in such a way that it can be used in classification tasks, such as named entity recognition. This kind of usage differs from the usual application of grammatical inference algorithms, as the verification or the syntactic analysis of sentences according to a grammar is not required. Instead, we are interested mainly in recognising sentence parts (phrases) and their classification in predefined semantic categories. The evaluation of this new algorithm has been performed on both synthetic languages, as well as on real world data for the task of relation extraction between named entities.

2 Information Extraction

Information extraction (IE) is the task of automatically extracting structured information from unstructured documents, mainly natural language texts. Due to the ambiguity of the term "structured information", information extraction covers a broad range of research, from simple data extraction from Web pages using patterns and regular grammars, to the semantic analysis of language for extracting meaning, such as the research areas of word sense disambiguation or sentiment analysis. The basic idea behind information extraction (the concentration of important information from a document into a structured format, mainly in the form of a table) is fairly old, with early approaches appearing in the 1950s, where the applicability of information extraction was proposed by the Zellig Harris for sub-languages, with the first practical systems appearing at the end of the 1970s, such as Roger Schank's systems [2,3], which exported "scripts" from newspaper articles. The ease of evaluation of information extraction systems in comparison to other natural language processing technologies such as machine translation or summarisation, where evaluation is still an open research issue, made IE systems quite popular and led to the Message Understanding Conferences (MUC) [4] that redefined this research field. Information extraction can be decomposed into several sub-tasks:

– Linguistic preprocessing, responsible for tasks such as token/sentence identification, part-of-speech tagging, morphological analysis, etc.
– Named-entity recognition, where domain specific entities, such as names of persons, organisations, and locations, monetary/time expressions, etc., are identified.
– Co-reference resolution, where named entity names or other mentions (such as pronouns) that refer to the same entity are grouped/related.
– Template element filling, a task responsible for grouping all properties of a real object into a single template element that represents the real task or event.
– Template relation, a task responsible for identifying relations among template elements.
– Scenario template, a task where related template elements are combined into a template that represents an event.

This thesis has investigated the use of machine learning in three key sub-tasks of information extraction: part of speech tagging, named entity recognition, and relation extraction. Part of speech tagging is an important sub-task of linguistic preprocessing, named entity recognition is an essential subtask of information extraction, and relation extraction is the main activity behind template element filling, template relation and scenario template.

2.1 Part of speech tagging

The term "part of speech tagging" refers to the process of assigning a unique tag to every word in a document, in a way that the part of speech of each word can be identified from its tag. Several approaches regarding this task for the Greek language have been presented in the literature, including the approach of Dermatas and Kokkinakis [5], where Hidden Markov Models were used, achieving an accuracy of 95%, when trained on a corpus of 110.000 words. Orphanos and Christodoulakis [6] combined decision trees with a morphological lexicon, achieving a performance of 93-95% regarding disambiguation of ambiguous words according to the lexicon, and an accuracy of 82-88% for words unknown to the lexicon. Papageorgiou et. al. [7] employed transformation-based error-driven learning (TBED) combined with a morphological lexicon, achieving an accuracy of 96% when trained on a corpus of 356.000 words. Finally, Malakasiotis [8] used active learning, achieving an accuracy of 80%, when trained on a corpus of 15.300 words.

This thesis examined the applicability of transformation-based error-driven learning (TBED) [9] to the Greek language [10], and the combination of TBED with a morphological lexicon [11], [12]. The contribution of this thesis to the task of part of speech tagging is three-fold:

– A tag set that extends the Penn Tree Bank tag set, in order to include information about gender, number and verb tenses.
– The first publicly available part of speech tagger for the Greek language.

– The accuracy of the proposed method approaches 98% when combined with a morphological lexicon, which is the higher reported accuracy of part of speech tagging for the Greek language.

2.2 Named entity recognition

The task of named entity recognition and classification (NERC) tries to identify names of "entities" in documents, and classify identified names into predefined semantic categories, which usually vary according to the thematic domain. A typical NERC system can be decomposed into three major subtasks: linguistic pre-processing, a lexicon, and a grammar. Linguistic pre-processing relates to tasks similar to tokenisation, sentence splitting, part-of-speech tagging, etc. The lexicon includes domain specific information, usually in the form of lists of known named-entities (gazetteer lists). Finally, the grammar is responsible for recognising the entities that are either not in the lexicon or appear in more than one gazetteer lists (disambiguation). The manual adaptation of those two resources to a particular domain is a time-consuming and in some cases impossible process, due to the lack of experts.

This thesis examined the applicability of machine learning as a solution to the problems of domain adaptation and performance tunning, by examining the following cases:

– The substitution of the grammar sub-system with machine learning.
– The adaptation/enrichment of the lexicon.
– The detection of when a NERC system is outdated and needs to be adapted to the (possibly changed) domain.

Regarding the substitution of grammar with machine learning, two approaches have been studied. In the first approach various machine learning algorithms have been examined, including symbolic ones (such as decision trees [13] and sub-symbolic ones (such as feed-forward multi-layered perceptrons [14]), using a representation proposed by this thesis for representing variable-length named-entities as vectors of constant size. The algorithms were compared to existing manually constructed systems on two languages (the MITOS [15,16] system for the Greek language, and the VIE [17] system for the English language). The results for the best performing algorithm (C4.5) are shown in figure 1, compared to the results of the two manually constructed NERC systems for the two languages. The results show that the proposed approach outperforms the VIE system for English, but does not outperform the MITOS system for the Greek language. In addition, experiments showed that maintaining word order is not important, as representations that ignore word order achieve comparable or better performance than representations that maintain it.

The second approach examines the *combination* of machine learning algorithms, which exploit different kinds of input information. A set of classification algorithms that try to detect whether a word is part (or not) of a named entity, are combined through a majority voter in order to classify all words in a

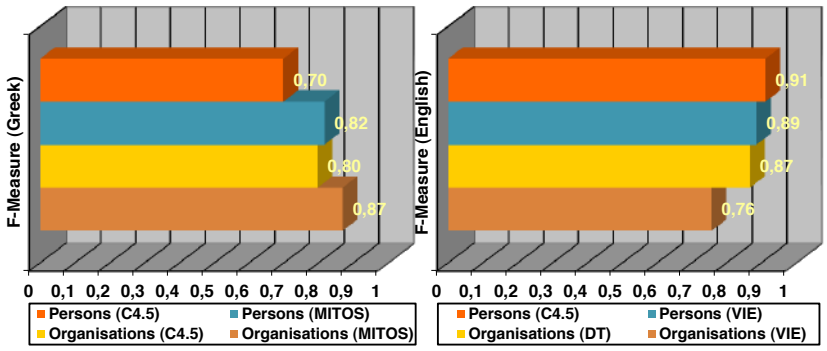


Fig. 1. Evaluation results of approach A, for the Greek and English languages.

document. In addition, noun phrases are identified, and classified as named entities through decision trees. The architecture of the second approach is shown in figure 2, where the evaluation results for Greek and English are shown in figure 3. Figure 4 shows the evaluation results of a NERC system developed by University of Edinburgh, UK, which is a hybrid system relying on manually constructed rules and machine learning [18,19] that has been trained and evaluated on the same English corpus as the ML-HNERC system (approach B). The results show that performance is higher for the Greek language compared to English, satisfying the objective for this system, which was motivated by the lower performance of approach A in Greek, compared to the English language. In addition, the performance of approach B on the English language is comparable to the performance of the system build by University of Edinburgh, which also build the top-scoring NERC system in the MUC-7 [20] competition. Maintaining the performance of a NERC system as a thematic domain timely evolves was also studied. The proposed approach makes an innovative use of machine learning, not to perform a task but rather to monitor the performance of another system. A machine learning based system (controller) is trained on the results of the system that will be monitored, and the deviation between the results of the two systems (the monitored and the control one) is recorded. As time passes and the thematic domain changes, the deviation between the two system is expected to change, as the two systems produce results based on possibly different input information. When the deviation exceeds a manually configured threshold, it is an indication that the monitored system is outdated, and needs to be updated. More details can be found in the thesis as well as in [21].

2.3 Relation extraction between recognised named entities

Relation extraction is the task of identifying semantic relations that hold between interesting entities in text data, and classify them into proper semantic categories. Being a challenging subtask of information extraction, it extracts the knowledge required to move from named entity recognition to data interpreta-

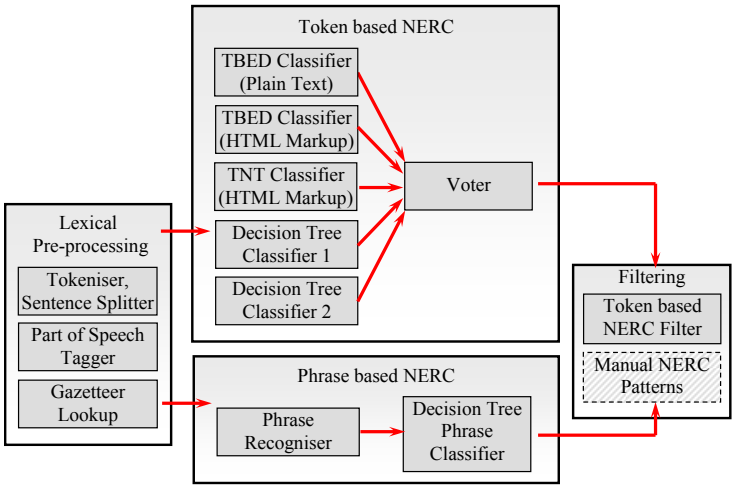


Fig. 2. The architecture of the ML-HNERC system (approach B).

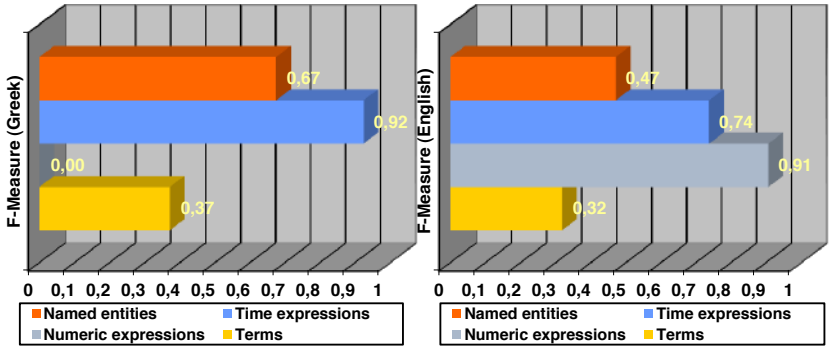


Fig. 3. Evaluation results of approach B, for the Greek and English languages.

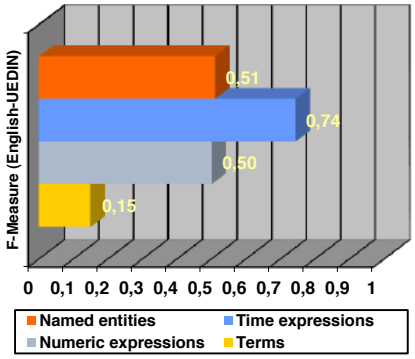


Fig. 4. Evaluation results of the system developed by University of Edinburgh, UK.

tion and understanding. The motivation behind the proposed approach is the simplification of the representations used in order to apply machine learning on natural language processing tasks, while the main objective is to examine the suitability of *grammatical inference* for the task of relation extraction. In this thesis, a supervised machine learning approach is proposed: assuming the existence of a named entity recogniser (NERC), the proposed approach extracts binary relations between named entities already identified in texts. Operating at the sentence level, a context-free grammar (CFG), which captures the patterns connecting related entities, is inferred from positive examples only. A new grammatical inference algorithm, eg-GRIDS+ ([22,23]), has been developed in order to infer a CFG only from positive examples. The need for negative feedback to control overgeneralisation, is eliminated through the use of minimum description length (MDL) [24].

The eg-GRIDS+ algorithm: A bias towards "simple" grammars As eg-GRIDS+ uses no negative evidence, an additional criterion is needed to direct the search through the space of context-free grammars and avoid overly general grammars. The approach of *minimum description length (MDL)* has been adopted in eg-GRIDS+, which directs the search process towards grammars that are compact, i.e., ones that require few bits to be encoded, while at the same time they encode the example set in a compact way, i.e. few bits are required to encode the examples using the grammar. Assuming a context-free grammar G and a set of examples (sentences) T that can be recognised (parsed) by the grammar G, the total description length of a grammar, henceforth *model description length* abbreviated as ML, is the sum of two independent lengths:

- The grammar description length (GDL), i.e. the bits required to encode the grammar rules and transmit them to a recipient who has minimal knowledge of the grammar representation, and
- The derivations description length (DDL), i.e. the bits required to encode and transmit all examples in the set T as encoded by grammar G, provided that the recipient already knows G.

The first component of the ML directs the search away from the sort of trivial grammar that has a separate rule for each training sentence, as this grammar will have a large GDL. However, the same component leads to another sort of trivial grammar, a grammar that accepts all sentences. In order to avoid this, the second component estimates the *derivation power* of the grammar, by measuring the way the *training examples* are generated by the grammar, and helps to avoid overgeneralisation by penalising general grammars. The higher the derivation power of the language, the higher its DDL is expected to be. The initial overly specific grammar is trivially best in terms of DDL, as usually there is a one-to-one correspondence between the examples and the grammar rules, i.e. its derivation power is low. On the other hand, the most general grammar has the worst score, as it involves several rules in the derivation of a single sentence, requiring substantial effort to track all the rules involved in the generation of the sentence.

Architecture of eg-GRIDS+ and the learning operators eg-GRIDS+ uses the training sentences in order to construct an *initial, "flat" grammar*. Then eg-GRIDS+ generalises this initial grammar, using one of the two available iterative search processes: beam or genetic search. Both search strategies utilise the same search operators in order to produce more general grammars:

- Merge NT: merges two non-terminal symbols into a single symbol, thereby replacing all their occurrences in all rules with the new symbol.
- Create NT: creates a new non-terminal symbol X, which is defined as a sequence of two or more existing non-terminal symbols. X is defined as a new production rule that decomposes X into its constituent symbols.
- Create Optional NT: duplicates a rule created by the "Create NT" operator and appends an existing non-terminal symbol at the end of the body of the rule, thus making this symbol optional.
- Detect Center Embedding: aims to capture the center embedding phenomenon. This operator tries to locate the most frequent four-gram of the form "A A B B", which is replaced by a new non-terminal symbol X and a new rule of the form "$X \rightarrow A\,A\,X\,B\,B$". All symbol sequences that match the pattern "$A+\ X?\,B+$" are replaced with X.
- Rule Body Substitution: examines whether the body of a production rule R is contained in bodies of other production rules. In such a case, every occurrence of the body of rule R in other rule bodies is replaced by the head of rule R.

Evaluation For the purposes of the evaluation, annotated corpus from the BOEMIE research project was used, which contained 800 HTML pages, retrieved from various sites of athletics associations, containing pages with news, results and athlete's biographies. Evaluation was performed through 10 fold-cross validation, and performance was measured in terms of precision, recall and F-measure. The obtained performance results are shown in table 1. Evaluation

	egGRIDS+			CRF++		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Name-Ranking	95.05 %	54.07 %	68.57 %	77.40 %	60.47 %	67.80 %
Name-Performance	92.14 %	49.26 %	64.17 %	84.42 %	84.18 %	84.93 %
Name-Country	98.85 %	88.88 %	93.58 %	88.78 %	84.63 %	86.70 %
Name-Gender	99.21 %	79.17 %	88.00 %	65.22 %	36.78 %	42.43 %
Name-Age	100.00 %	98.11 %	99.04 %	79.88 %	56.03 %	64.28 %
Overall	**96.48 %**	**65.96 %**	**78.32 %**	**79.88 %**	**60.47 %**	**67.80 %**

Table 1. Relation extraction performance results.

results suggest that the proposed approach performs well in comparison to the state of the art, despite the difficulties of comparing results obtained on different

corpora. For example, in [25], the presented approach, expanding on a basis of 55 manually constructed seed rules, exhibits precision around 88% with 43% recall on 1032 news reports on Nobel prizes from New York Times, BBC and CNN. In addition, Conditional Random Fields (CRF++) were applied on the same corpus, achieving lower results than the approach based on eg-GRIDS+.

3 Conclusions

This thesis proposes the exploitation of machine learning in nodal points of a typical information extraction system, having as aim the assistance adapting the system into new thematic domains and perhaps languages. This first research topic of this thesis involves part of speech tagging for the Greek language. Transformation-based error-driven learning (TBED) has been applied for the first time in the Greek language, achieving high performance, directly comparable with corresponding systems for the Greek language, requiring at the same time considerably less training data. Simultaneously, the approach that is described in this thesis constituted the first Greek part-of-speech tagger that has been distributed freely, as an open source application, with important acceptance from the scientific community, as denoted by the number of citations in the relative publications.

The second research topic concerns the area of named entity recognition. Three machine learning algorithms were examined for this task, both symbolic and stochastic ones, achieving satisfactory results. The algorithms were examined in various thematic domains, in both the English and Greek languages, confirming not only the ability of machine learning to support the task of named entity recognition, but also the adaptability of the machine learning based approaches not only in new thematic domains, but also in languages. The research that has been contacted in the context of this thesis is included among the first information extraction systems for the Greek language that have been reported in the bibliography. In addition, it has been observed that, at least for the task of named entity recognition, the order of words in a sentence is not important. Despite the fact that initially this observation seemed surprising, the widespread use of the "bag-of-words" representation - which also ignores the word order - not only for named entity recognition, but also for other natural language processing tasks, shows the correctness of this initial observation.

The third research topic concerns the development of new machine learning algorithm, able to infer context free grammars from positive only examples. An important characteristic of this new algorithm is its ability to processes large volumes of data, a consequence of the observation that it is computationally cheaper to predict the result of applying a learning operator, than to apply the operator and evaluate the produced grammar. The results achieved by the new algorithm on the "Omphalos" [26] competition were also significant, where it solved the first problem without human intervention within the competition time period, while it has been successfully combined with the winning algorithm,

removing the need of the winning algorithm for human intervention, in cases where its heuristic could not drive further the learning process.

References

1. Langley, P., Stromsten, S.: Learning context-free grammars with a simplicity bias. In: Proceedings of the 11th European Conference on Machine Learning. ECML '00, London, UK, UK, Springer-Verlag (2000) 220–228
2. Schank, R.C., Abelson, R.P.: Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures. L. Erlbaum, Hillsdale, NJ (1977)
3. Schank, R.C., Kolodner, J.L., DeJong, G.: Conceptual information retrieval. In: Proceedings of the 3rd annual ACM conference on Research and development in information retrieval (SIGIR '80), Cambridge, UK (1980) 94–116
4. Marsh, E., Perzanowski, D.: Muc-7 evaluation of ie technology: Overview of results. In: Proceedings of the Seventh Message Understanding Conference (MUC-7), `http://www.itl.nist.gov/iaui/894.02/related_projects/muc/index.html` (1998)
5. Dermatas, E., Kokkinakis, G.K.: Automatic stochastic tagging of natural language texts. Computational Linguistics **21**(2) (1995) 137–163
6. Orphanos, G.S., Christodoulakis, D.N.: Pos disambiguation and unknown word guessing with decision trees. Computer Engineering (1999) 134–141
7. Papageorgiou, H., Prokopidis, P., Giouli, V., Piperidis, S.: A unified pos tagging architecture and its application to greek. In: Proceedings of the 2nd Language Resources and Evaluation Conference, Athens, European Language Resources Association (June 2000) 1455–1462
8. Malakasiotis, P.: Αναγνώριση μερών του λόγου σε ελληνικά κείμενα με τεχνικές ενεργητικής μάθησης (Part-of-speech tagging in Greek texts using active learning techniques). Master's thesis, Department of Informatics, Athens University of Economics and Business (2005)
9. Brill, E.: Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. Comput. Linguist. **21**(4) (December 1995) 543–565
10. Petasis, G., Paliouras, G., Karkaletsis, V., Spyropoulos, C.D.: Resolving Part-of-Speech Ambiguity in the Greek Language Using Learning Techniques. In: Proceedings of the ECCAI Advanced Course on Artificial Intelligence (ACAI '99), Chania, Greece (July 5–16 1999)
11. Petasis, G., Karkaletsis, V., Farmakiotou, D., Androutsopoulos, I., Spyropoulos, C.D. In: A Greek Morphological Lexicon and Its Exploitation by Natural Language Processing Applications. Volume 2563 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2003) 401–419 http://www.springerlink.com/content/hcdjrlvj5nlybf5c/.
12. Spiliotopoulos, D., Petasis, G., Kouroupetroglou, G.: Prosodically Enriched Text Annotation for High Quality Speech Synthesis. In: Proceedings of the 10th International Conference on Speech and Computer (SPECOM-2005), Patras, Greece (October 17–19 2005) 313–316
13. Quinlan, J.R.: C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1993)
14. Perantonis, S.J., Ampazis, N., Virvilis, V.: A learning framework for neural networks using constrained optimization methods. Annals of Operations Research **99**(1) (2000) 385–401

15. Petasis, G., Petridis, S., Paliouras, G., Karkaletsis, V., Perantonis, S.J., Spyropoulos, C.D.: Symbolic and Neural Learning for Named-Entity Recognition. In: Proceedings of European Best Practice Workshops and Symposium on Computational Intelligence and Learning (COIL 2000), Chios, Greece (June 19–23 2000) 58–66
16. Petasis, G., Petridis, S., Paliouras, G., Karkaletsis, V., Perantonis, S.J., Spyropoulos, C.D. In: Symbolic and Neural Learning of Named-Entity Recognition and Classification Systems in Two Languages. Volume 18 of International Series in Intelligent Technologies. Springer Berlin / Heidelberg (January 2002) 193–210 http://www.springer.com/mathematics/book/978-0-7923-7645-3.
17. Humphreys, K., Gaizauskas, R., Cunningham, H., Azzam, S.: GATE: VIE technical specifications. Technical report, ILASH, University of Sheffield (1997) Included in the documentation of GATE 1.0.0.
18. Curran, J.R., Clark, S.: Investigating gis and smoothing for maximum entropy taggers. In: Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - Volume 1. EACL '03, Stroudsburg, PA, USA, Association for Computational Linguistics (2003) 91–98
19. Curran, J.R., Clark, S.: Language independent ner using a maximum entropy tagger. In: Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4. CONLL '03, Stroudsburg, PA, USA, Association for Computational Linguistics (2003) 164–167
20. Chinchor, N.A.: Proceedings of the Seventh Message Understanding Conference (MUC-7) named entity task definition. In: Proceedings of the Seventh Message Understanding Conference (MUC-7), Fairfax, VA (April 1998) 21 pages version 3.5, http://www.itl.nist.gov/iaui/894.02/related_projects/muc/.
21. Petasis, G., Vichot, F., Wolinski, F., Paliouras, G., Karkaletsis, V., Spyropoulos, C.D.: Using Machine Learning to Maintain Rule-based Named - Entity Recognition and Classification Systems. In: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. ACL '01, Toulouse, France, Association for Computational Linguistics (July 9–11 2001) 426–433
22. Petasis, G., Paliouras, G., Karkaletsis, V., Halatsis, C., Spyropoulos, C.D.: E-GRIDS: Computationally Efficient Grammatical Inference from Positive Examples. GRAMMARS **7** (2004) 69–110 Technical Report referenced in the paper: http://www.ellogon.org/petasis/bibliography/GRAMMARS/GRAMMARS2004-SpecialIssue-Petasis-TechnicalReport.pdf.
23. Petasis, G., Paliouras, G., Spyropoulos, C.D., Halatsis, C.: Eg-GRIDS: Context-Free Grammatical Inference from Positive Examples Using Genetic Search. In Paliouras, G., Sakakibara, Y., eds.: Grammatical Inference: Algorithms and Applications, Proceedings of the 7th International Colloquium on Grammatical Inference (ICGI 2004). Volume 3264 of Lecture Notes in Computer Science., Athens, Greece, Springer Berlin / Heidelberg (October 11–13 2004) 223–234
24. Rissanen, J.: Stochastic Complexity in Statistical Inquiry Theory. World Scientific Publishing Co., Inc., River Edge, NJ, USA (1989)
25. Xu, F., Uszkoreit, H., Li, H.: A seed-driven bottom-up machine learning framework for extracting relations of various complexity. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, Association for Computational Linguistics (June 2007) 584–591
26. Starkie, B., Coste, F., van Zaanen, M.: The omphalos context-free grammar learning competition. In Paliouras, G., Sakakibara, Y., eds.: Grammatical Inference: Algorithms and Applications, Proceedings of the 7^{th} International Colloquium, ICGI 2004, Athens, Greece, October 11-13, 2004. Volume 3264 of Lecture Notes in Computer Science., Springer (2004) 16–27

A Software Environment for the Development of Component-based Augmentative and Alternative Communication Applications

Alexandros Pino*

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
pino@di.uoa.gr

Abstract. As an answer to the disabled community's Odyssey to gain access to adaptable, modular, multilingual, cheap and sustainable Augmentative and Alternative Communication (AAC) products, we propose the use of the ITHACA framework. It is a software environment for building component-based AAC applications, grounded on the Design for All principles and a hybrid (community and commercial) Open Source development model. ITHACA addresses the developers, the integrators (e.g., vendors, facilitators, special educators), as well as people who use AAC. We introduce a new AAC product lifecycle, i.e., the design-develop-distribute procedures, from the developers' viewpoint and the search-select-modify-maintain procedures from the integrators' perspective. ITHACA provides programmers with a set of tools and reusable Open Source code for building AAC software components. It also facilitates AAC product integrators to put together sophisticated applications using the available on the Web, independently pre-manufactured, free or commercial software parts. Furthermore, it provides people who use AAC with a variety of compatible AAC software products which incorporate multimodal, user-tailored interfaces that can fulfill their changing needs. Several ready to use ITHACA-based components, including on-screen keyboards, Text-to-Speech, symbol selection sets, e-chatting, e-mailing, and scanning-based input, as well as four complete communication aids addressing different user cases have been developed. This demonstration showed good acceptance of the ITHACA applications and substantial improvement of the end users' communication skills. Developers' experience on working in ITHACA's Open Source projects was also positively evaluated.

Keywords. Augmentative and Alternative Communication, Open Source, Component-Based Development, Design for All, Symbolic Communication Systems

* Dissertation Advisor: Georgios Kouroupetroglou, Assoc. Professor

1 Introduction

1.1 The Problem

For people with complex communication needs, those with speech and/or motor impairments, cognitive limitations, learning disabilities and aging, daily routine as well as rehabilitation and educational programs often include the use of Augmentative and Alternative Communication (AAC) aids [1]. In the past, AAC was dominated by low-technology or non-electronic devices [2]. A few decades ago, several electronic aids with voice recording and playback capabilities were introduced in the international market. Such non-computer-based products are still widely used. Although these devices are considered very useful for the persons using AAC, they provide a limited vocabulary, need extra effort from the facilitators to add new recordings, and cannot keep up with the nontrivial progress usually achieved by their users [3].

Recently, several computer-based communication aids that support a range of symbolic communication systems, special I/O devices, configurable User Interfaces (UIs), and speech synthesis have been developed by various companies. These devices have larger vocabularies, but they support a very limited number of natural languages and it is rather impossible to add new ones. These modern computer-mediated interpersonal communication systems should be adaptable in order to satisfy the wide variety of the users' changing needs [4] and the specific user profiles. Nevertheless, traditionally, software application developers in the domain of communication aids are creating stand-alone, monolithic applications based on their studies of user needs and market research. Retailers do not get actively involved in the development or in the configuration and adaptation process of communication aids. The only possible feedback in the product life cycle is between the end user and the reseller and that feedback is difficult to propagate to the developer. Furthermore, AT products are very few and expensive due to the small market and the lack of software reuse, as many manufacturers develop the same functionalities and features from scratch again and again. Throughout each product's life cycle, from the original idea to the end user, there is no significant feedback and evaluation. Finally, finding the right product for specific user needs is a difficult task due to the dispersed information and selling points.

Other problems that disabled users face with existing computer-based commercial solutions include: absence of multilingual support, lack of proper support for customization, and difficulty in adding or removing functionalities or components when needed. Moreover, designing and developing interpersonal communication aids for people with disabilities is a domain for which modern software engineering approaches such as those that combine Component-Based Development (CBD) and the Open Source model that lower the development costs have not been applied. DeRuyter et al. [5] refer to the development of Open Source software that runs on mainstream computers, as a better alternative in order to provide maximum flexibility and accessibility. ITHACA framework is addressing these problems by combining the Design for All, Open Source, and CBD approaches.

1.2 Related work

Projects COMSPEC [6] and ACCESS [7], have made significant steps towards CBD. ComLink and ATIC were two component-based approaches produced by these projects as an answer to the problems of the AAC market. Although both these frameworks were characterized "open", which meant that third-party developers could theoretically develop compatible components, they were essentially Closed Source, as their code was not freely available. The AAC industry's response was not encouraging, as no third party components were delivered to enrich the basic component collection that accompanied these two frameworks.

The ULYSSES framework was the product of the subsequent project AENEAS [8-10]. The main difference from its predecessor was that ATIC used a proprietary Message Manager and a complex communication protocol between components, making the conformance with the specific architecture difficult for the developers. On the other hand, ULYSSES used a widely available and known infrastructure and messaging system that was embedded in the (Operating System) OS, and a simpler object model, making its guidelines and specifications straightforward. Nevertheless, ULYSSES had the main drawback that the AAC industry needed to be accustomed to its proprietary guidelines and code, in order to comply with the framework. This is very unlikely to happen, especially when the framework is closed-source.

The World Wide Augmentative and Alternative Communication (WWAAC) project has contributed towards the direction of Open Source development, in the domain of Internet accessibility for AAC users [11]. The most important contribution of the WWAAC project was the Concept Coding Framework (CCF) [12]. CCF provides direct support for symbol users on web pages through its open-sourced concept coding infrastructure and protocol. The vision of concept coding is that instead of images and symbols having to be transferred from one computer to another, it should be possible to transmit a unique code designating the meaning of the symbol needing to be transferred. Nevertheless, the WWAAC project did not develop a framework for building applications such as ComLink, ATIC, ULYSSES, and ITHACA, and its resulting applications were not compatible with any of these frameworks.

OATSoft [13] is an Open Source software repository dedicated to Assistive Technology (AT). It currently lists more than 150 Open Source projects and it is the most important initiative in this domain. Project:Possibility is a similar initiative that hosts more than 10 Open Source AT projects. Both initiatives have an active community formed, but neither hosts an Open Source component-based AAC development framework project.

1.3 The ITHACA-based AAC Software Lifecycle

The ATIC architecture [7] proposed a different life cycle that solved some of the aforementioned problems and introduced an extended role for communication aids resellers. They were considered as an important user group (namely the integrators) having an essential part in the life cycle of the developed products, with the task of

assembling the whole AT system from available software components and suitable I/O devices and techniques.

ULYSSES introduced the important role of the Internet as a widely accessible medium for gathering and propagating information about the framework and available software components and I/O devices. In ATIC, the stores that were specialized in AT products played the role of the component repository. ULYSSES replaced the traditional stores with a specialized website offering a higher degree of availability, variety and flexibility.

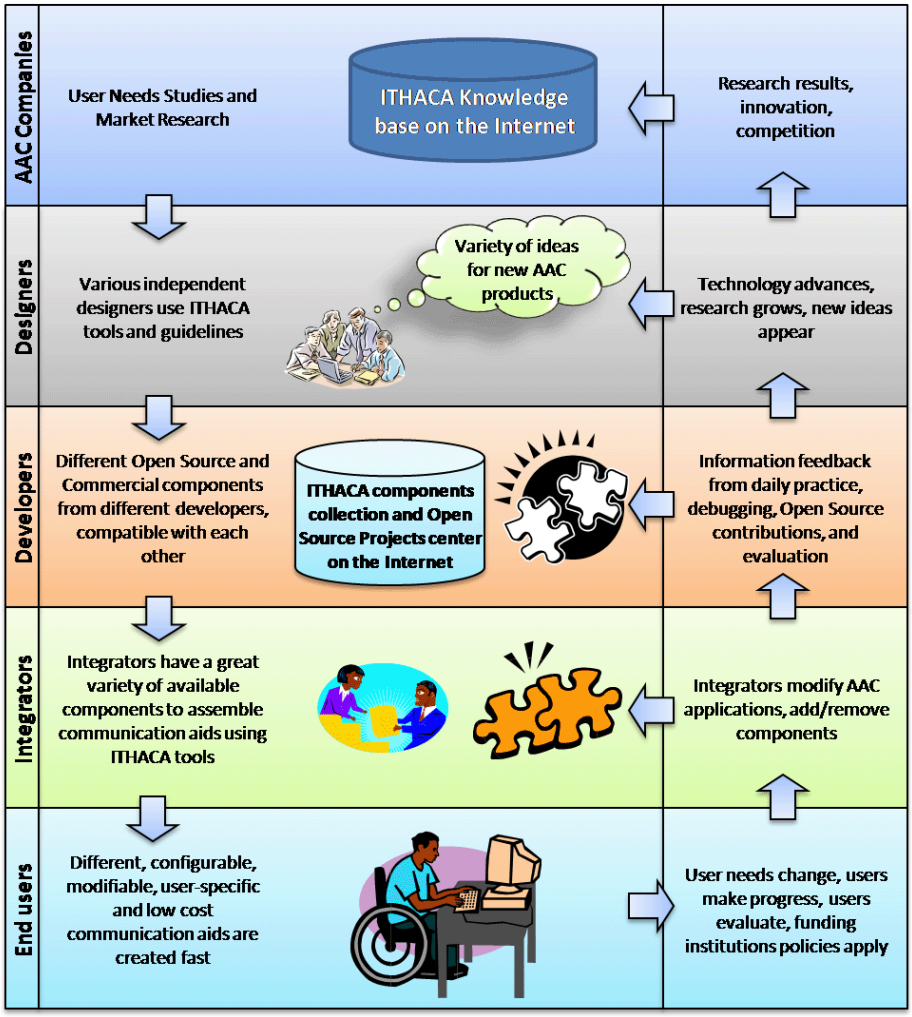


Fig. 1. ITHACA AAC product Life Cycle

With ITHACA, the decision as to which AAC device or software is purchased, how instruction is provided, and how the device is maintained and developed typically

involves many individuals, including the person who uses AAC, family members, communication and education professionals, and funding agencies [14]. All stakeholders included in the AAC product lifecycle can participate and contribute to Open Source software development, both on technical and non-technical aspects. Developers concentrate on the technical aspects of the framework regarding the Open Source software engineering techniques, interfaces, and guidelines. On the other hand, integrators focus on the proposed product life cycle, integration methods, and administrative tools for installing, configuring, modifying and maintaining the applications.

ITHACA proposes an extended and upgraded AAC product life cycle by introducing a Web-based AAC knowledge base and the Web-based component inventory (Figure 1). Most importantly, the online ITHACA components may be Open Source (community or commercial) or Closed Source, thus modifying the lifecycle to a new form of hybrid Open Source/Closed Source product lifecycle. A substantial aspect of the new life cycle is the high importance that is given to information propagation between all stakeholders and all stages.

The framework allows for a new viewpoint from the side of AAC production stakeholders. The following procedures are positively affected: *Design, Development,* and *Distribution* of AAC applications. From the end-users' ITHACA also allows for innovative approaches in lifecycle stages: *Search, Selection, Modification,* and *Maintenance* of AAC products.

2 The ITHACA Framework

What mainly sets ITHACA apart from the aforementioned projects is that it proposes a hybrid approach, by combining CBD, a free Open Source framework and core components, along with community and commercial Open Source peripheral components in order to increase competition, while maintaining the interests of software manufacturers for this market. Universal Design or Design for All completes the framework, by offering the important aspects of adaptability and multiple users' needs inclusion [15]. Two main user groups of the ITHACA framework are identified: the AAC manufacturers or software developers, and the AAC systems integrators or communication aids resellers. Although both user groups are aware of the basic characteristics of the framework, each must know different aspects of ITHACA in detail.

2.1 Technological Background

ULYSSES – the framework on which ITHACA is based – proposed the use of a combination of the following specifications, models and services for the development of components and software applications [9]: Application Specification for Microsoft Windows 2000 for Desktop Applications, Component Object Model (COM) specification, and COM's Extension for Component Services (COM+) .

ITHACA proposes an update of ULYSSES specifications and guidelines by replacing Application Specification for Microsoft Windows 2000 for Desktop Applications with the contemporary Microsoft UI Automation for upgraded application accessibil-

ity. This Microsoft specification was chosen against its older version Microsoft Active Accessibility and IBM's IAccessible2 as the best practice for designing and developing Windows-based accessible applications. ITHACA also adopted COM and COM+ specification as they were updated and supported by the .NET programming environment.

The PC with Microsoft Windows platform was chosen as the OS on which ITHACA-based communication aids run, because of the advanced accessibility options it provides, the user-friendliness, the increased system stability, and the high support for special I/O devices. Due to the large installed basis and availability of the OS, users will not need to buy a new computer or have to install a non-common or hard to use OS to use ITHACA-based communication aids.

ITHACA framework makes extensive use of COM+ Events [16] and the corresponding model, which is an evolution of the client-server model. COM+ was chosen as the basis of the framework's architecture because its services are widely used and accepted by the developer's community. One of the most important features that COM+ provides is the Component Management Console, which is a powerful tool for managing and maintaining COM+ applications; this tool clears many technical problems and restrictions previous frameworks had.

The ITHACA framework produces .NET applications that use COM+ services, and provides a specific communication protocol between software components. This protocol is open and easily modified according to the application's needs for data exchange between its components. The protocol is based on the consideration that the fundamental data used in AAC are the "Concepts" (an idea widely accepted in the AAC domain). Abstract/logical concepts can engage various data types at the presentation level; that is, concepts may be represented by strings (i.e., words or phrases), video, sound, or icons. A concept that is conveyed from one component to another, locally or remotely, may be processed and may change data type and/or language between components, making its management rather complex.

To simplify the situation, we defined a base language (database of concepts) named "Interlingua", i.e. a pseudo-language based on English, in which all concepts can be represented as character strings. Language-independent components can communicate using Interlingua, while output components or language-aware (natural or symbolic) components use the equivalent representation (in any form) of the Interlingua concept in their own language or symbolic system according to the database's relations. The database of natural and symbolic languages (including Interlingua, photograph, sound or video-based languages) is the heart of the framework and allows simplification of inter-component communication by using strings only. Final users and/or facilitators and educators can add new content to the database, and add or modify concepts of the Interlingua.

The ITHACA concept transmission protocol consists of a set of interfaces used as a common channel for propagating strings of data, i.e., characters, words, sentences or complete messages that the user composes. This communication protocol and interfaces, in cooperation with the ITHACA database and the Interlingua concepts helps to overcome translation and inter-component communication related challenges that AAC component-based framework research has encountered for many years.

2.2 ITHACA for Developers

ITHACA guides developers of AAC components to program their software modules as Publishers or Subscribers to data provided through the COM+ Event Service. A developer, who uses the corresponding Event Classes following straightforward guidelines, can create components incorporating their own GUI, Speech User Interface (SUI) or plain transparent functionality, which can interoperate with other ITHACA-compliant components that may have been created by other developers.

An executable program, also provided with ITHACA, initializes and configures the AAC application when run for the first time, and activates all components and interfaces every subsequent time that it is executed. The framework's executable program, all DLLs, along with their Open Source code and guidelines, can be downloaded from ITHACA's website[1].

Furthermore, ITHACA provides ready-to-use software components to AAC developers for testing the communication of their components with the Event Service and other components. These test or "template" components serve as Publishers, Subscribers or both Publishers and Subscribers of data. Test components are also Open Source offering basic framework code examples, and can create a real application environment in order to verify the correct operation of the component being tested.

Finally, ITHACA framework is intrinsically Internet-ready, i.e., it provides all the necessary infrastructure and support for components that implement remote synchronous (chat) or asynchronous (e-mail) interpersonal communication using available Internet technologies and data transfer protocols [17].

2.3 ITHACA for Integrators

An integrator can easily assemble and manage communication aids from various independently developed components, which cooperate to provide the application functionality and UI. For integrators of interpersonal communication aids, ITHACA offers a detailed user guide and an installation program as well as a World Wide Web information center with a catalog of ITHACA-compatible components (free or commercial) to choose from. Furthermore, the database of symbolic communication systems and natural languages is also available to integrators, and ready to be connected with ITHACA components to provide concept mapping for interpersonal communication applications. For a specific AAC application, only parts of the database are needed and can be downloaded separately.

2.4 ITHACA Components

The most common functionality in AAC products has already been developed using the ULYSSES framework, in the form of pluggable components [9]. All these components are reprogrammed, converted to ITHACA-compliant .NET modules, Open Source and freely available on the ITHACA website. Based on Open Source software

[1] http://speech.di.uoa.gr/ithaca/

development projects, software companies and hobbyist programmers including special education professionals contributed to the final result. The code reuse was extensive and this allowed for better UI refinement and better interoperability between all modules. The main components we developed are: *Virtual Keyboard, Word Selection Set, Symbol Selection Set, Symbol or Text Editor, Scanning, Syntactic parser* [18], *Text-to-Speech System* [19], *Chat, E-mail, Database/Translation* [17].

3 Framework's Evaluation

For a nine months period, 20 programmers and researchers were involved in the process of designing and developing ten framework core components and eleven ITHACA compliant AAC components. The developers' team consisted of 6 pregraduate, 5 postgraduate and 4 doctoral students of the Information and Telecommunications Department of the University of Athens. Additionally, 2 professional developers (employees of a major software company) also participated in the development team. The project leaders' team consisted of 3 researchers who were previously involved in ATIC and ULYSSES, and currently in ITHACA project. This team also played the frameworks' designers', core components' developers', and integrators' role. Forum and online chatting were used for inter- and intra-team communication. The programming environment was .NET-based, and Visual Basic, C++ and C# programming languages were used.

Table 1. Developers' rating results

Measure	Rating	Indicators	mean	std dev
Functionality	**4.39**	*Suitability*	4.82	0.39
		Accurateness	4.35	0.61
		Interoperability	4.00	0.87
Maintainability	**4.09**	*Required support*	3.24	0.66
		Changeability	4.65	0.49
		Stability	3.65	0.61
		Testability	4.82	0.39
Reliability	**3.73**	*Maturity*	3.59	0.80
		Fault tolerance	3.53	0.80
		Recoverability	4.06	0.83
Portability	**4.35**	*Adaptability*	4.65	0.49
		Installability	3.59	1.18
		Replaceability	4.82	0.39
Usability	**4.31**	*Undestandability*	4.65	0.61
		Learnability	4.71	0.47
		Operability	3.59	1.12

Several models in the literature for evaluating Open Source projects [20], Open Source software quality [21], as well as Open Source repositories success, were con-

sidered for the selection of the aspects of ITHACA that were evaluated. A combination of features and measures was used for a subjective evaluation of the quality framework by the 17 participating developers, using an online survey with 5 measured domains and a 1-5 Likert rating scale. The scale was defined as: 1=very bad; 2=bad; 3=moderate; 4=good; and 5=very good. A total number of 16 indicators were rated. The ISO 9126 standard that classifies software quality in a structured set of characteristics was taken as a reference, though some modifications were made to the set of sub-characteristics that the standard suggests in order to better suit an Open Source framework evaluation as opposed to a software application evaluation. Opinions were also asked in the form of open type questions, investigating ITHACA's functionality, maintainability, reliability, portability, and usability. The most important findings are summarized in Table 1.

In order to confirm the breadth of software that can be produced as a proof of concept of the ITHACA framework, we have further conducted a number of demonstrations with real users. Making combinations of the aforementioned components, we assembled a range of customized AAC applications addressing various user needs and communication requirements. We briefly summarize our observations regarding the experiences of the users, their family members, and their teachers. Our goal is not to provide a formal evaluation of the resulting systems, but to provide a sense as to how the systems were received by the individuals for whom they were designed.

Fig. 2. The four demonstration AAC applications

Four users were selected from a special education and rehabilitation center in Greece; they were all children monitored in their special school environment, and the appropriate communication aids were selected following the main guidelines proposed in [22]. All participants were diagnosed as having cerebral palsy, with different symptoms that ranged from mild to very severe. All users had severe speech, and none of the users had ever used computer-based AAC before. The aim was to help all four users communicate in their Greek-speaking environment, firstly at school and secondly at home. Screenshots of the final GUIs of the four applications are illustrated in Figure 2.

4 CONCLUSIONS

We have presented the integrated ITHACA approach for the development of computer-based AAC applications. ITHACA consists of an Open Source, component-based framework that aims to simplify the integration of multi-vendor components into low cost AAC products and maximizes modularity and reusability [23]. ITHACA suggests technical (CBD), as well as business and management (Open Source, combined free and commercial) models for AAC assistive technology support provision. Following the Design for All approach, developers can build reliable components, adaptable to various user needs and requirements.

The mixed model that includes an Open Source framework, running on a mainstream proprietary OS and a mainstream computer, changes the life cycle of AAC products. It allows for Open Source, Closed Source, free, and commercial components, to compete for a place in the end-users application. Community volunteer work in an Open Source project context could be part of the solution for the high costs of AAC products. This way, the cost of debugging or even modifying software components remains low also, thus making both the development process and the final applications more sustainable [25]. Furthermore, AAC researchers have the opportunity to easily test their novel ideas and technology in an Open Source, component-based integrated environment without having to develop additional infrastructure.

ITHACA is a result of several research projects and the entire life cycle we introduced could be realized only if all stakeholders embraced this framework or a similar one. This could only be achieved with joined efforts by companies, institutes, funding organizations and people who use AAC. An important aspect for the success of this life cycle is the continuous flow of feedback information from all stakeholders and stages of the life cycle in all directions. The modified production and maintenance procedures should be coupled with a business model and a central administration system in order to be complete. A hybrid Open Source/Commercial model of management is proposed, meaning that an Organization (profit or non-profit) should coordinate development and distribution procedures. It is expected that the adoption of such an approach by the AAC industry and community will lead to affordable AAC products and upgrade their quality and variety.

The developers' evaluation showed that the potential contribution of the framework to the AAC domain is considered very important. The extensive code reuse, as

well as the good functionality of the framework, was highly appreciated. To reach a critical mass of basic components and functionality and to attract enough interest and more development resources is a concern for all Open Source initiatives, and the success of their dissemination is crucial. The listing of ITHACA in OATSoft and Project:Possibility websites, or SourceForge , the largest hub for Open Source development projects, will be our next step in order to disseminate the framework as an Open Source project.

Multiple combinations of ITHACA-compliant Open Source components, implementing various functionalities and UIs for interpersonal communication applications have been used to assemble four different applications utilizing different accessibility options as well as I/O devices and interaction techniques, revealing the flexibility of the model used. The demonstration of the four systems and real users' and their facilitators' comments showed that interacting computer-based AAC is always a new and interesting way of learning and communicating. Progress was made even after more traditional methods had reached their limits. Involving users and facilitators in the development, testing, debugging, and evolving procedures was considered very valuable and effective.

5 References

1. Beukelman, D.R., Mirenda, P.: Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults (3rd edition). Paul H. Brookes Publishing Co., Baltimore (2005)
2. Reichle, J., York, J., Sigafoos, J.: Implementing Augmentative and Alternative Communication: Strategies for Learners with Severe Disabilities. Paul H. Brookes Publishing Co., Baltimore (1991)
3. Cumley, G.D., Swanson, S.: Augmentative and Alternative Communication Options for Children with Developmental Apraxia of Speech: Three Case Studies. Augmentative and Alternative Communication 15, 110-125 (1999)
4. Light, J.: Toward a Definition of Communicative Competence for Individuals Using Augmentative and Alternative Communication Systems. Augmentative and Alternative Communication 5, 137–144 (1989)
5. Deruyter, F., McNaughton, D., Caves, K., Bryen, D.N., Williams, M.B.: Enhancing AAC Connections with the World. Augmentative and Alternative Communication 23, 3, 258-270 (2007)
6. Lundälv, M., Hekstra, D., Stav. E.: Comspec - a Java based Development Environment for Communication Aids. In: Placencia Porrero, I., Ballabio, E. (eds.) Improving the Quality of Life for the European Citizen, pp. 203-207. IOS Press, Amsterdam, (1998)
7. Kouroupetroglou, G., Viglas, C., Stamatis, C. Pentaris, F.: Towards the Next Generation of Computer-based Interpersonal Communication Aids. In: 4th European Conference for the Advancement of Assistive Technology, AAATE97, pp. 110-114. IOS Press, Amsterdam (1997)
8. Kouroupetroglou, G., Pino, A.: ULYSSES: A Framework for Incorporating Multi-Vendor Components in Interpersonal Communication Applications. In: 6th European Conference for the Advancement of Assistive Technology, AAATE 2001, pp. 55-59. IOS Press, Amsterdam (2001)

9. Kouroupetroglou, G., Pino, A.: A New Generation of Communication Aids under the ULYSSES Component-based Framework. In: 5th International ACM SIGCAPH Conference on Assistive Technologies, ASSETS 2002, pp. 218-225. ACM Press, New York (2002)
10. Kouroupetroglou, G., Pino, A., Viglas, C.: Managing Accessible User Interfaces of Multivendor Components under the ULYSSES Framework for Interpersonal Communication Applications. In: International Conference on Human-Computer Interaction, HCI 2001, pp. 185-189. Lawrence Erlbaum Associates, Inc., Mahwah (2001)
11. Poulson, D., Nicolle, C.: Making the Internet Accessible for People with Cognitive and Communication Impairments. Universal Access in the Information Society 3, 1, 48-56 (2004)
12. Judson, A., Hine, N., Lundälv, M., Farre, B.: Empowering Disabled Users through the Semantic Web. In: 1st International Conference on Web Information Systems and Technologies, WEBIST, pp. 162-167. INSTICC Press, Portugal (2005)
13. Judge, S., Lysley, A., Walsh, J., Judson A., Druce S.: OATS - Open Source Assistive Technology Software - a Way Forward. In: 12th Biennial Conference of the International Society for Augmentative and Alternative Communication (on CD-ROM). ISAAC (2006)
14. Rackensperger, T., Mcnaughton, D., Krezman, C., Williams, M., D'Silva, K.: When I First Got it, I Wanted to Throw it over a Cliff: The Challenges and Benefits of Learning Technology as Described by Individuals who use AAC. Augmentative and Alternative Communication 21, 165-186 (2005)
15. Savidis, A., Stephanidis, C.: Inclusive development: Software Engineering Requirements for Universally Accessible Interactions. Interacting with Computers 18, 1, 71-116 (2006)
16. Platt, D.: The COM+ Event Service Eases the Pain of publishing and Subscribing to Data. Microsoft Systems Journal 14, 9 (1999)
17. Viglas, C., Kouroupetroglou, G.: An open machine translation system for augmentative and alternative communication. In: Miesenberger, K., Klaus, J., Zagler, W.L. (eds.) ICCHP 2002. LNCS 2398, 698-706. Springer, Heidelberg (2002)
18. Karberis, G., Kouroupetroglou, G.: Transforming Spontaneous Telegraphic Language to Well-formed Greek Sentences for Alternative and Augmentative Communication. In: Vlahavas, I.P., Spyropoulos, C.D. (eds.) SETN 2002. LNCS 2308, 155-156. Springer, Heidelberg (2002)
19. Xydas, G., Kouroupetroglou, G.: The DEMOSTHeNES Speech Composer. In: 4th ISCA Tutorial and Research Workshop on Speech Synthesis, pp. 167-172. Kluwer Academic Publishers, Dordrecht (2001)
20. Crowston, K., Howison, H., Annabi, J.: Information Systems Success in Free and Open Source Software Development: Theory and Measures. Software Process Improvement and Practice 11, 2, 123-148 (2006)
21. Stamelos, I., Angelis, L., Oikonomou, A., Bleris, G.: Code Quality Analysis in Open-Source Software Development. Information Systems Journal, 2nd Special Issue on Open-Source, 12, 1, 43-60 (2002)
22. Woltosz, W.S.: A Proposed Model for Augmentative and Alternative Communication Evaluation and System Selection. Augmentative and Alternative Communication 4, 233-235 (1987)
23. Spinellis, D., Szyperski, C.: How Is Open Source Affecting Software Development? IEEE Software 21, 1, 28-33 (2004)
24. Tate, K.: Sustainable Software Development: An Agile Perspective. Addison-Wesley Professional, Boston (2005)

Virtualizing Information Spaces for the Expansion and Integration of Heterogeneous Data Collections and Systems

Kostas Saidis*
saiko@di.uoa.gr

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications

Abstract. The amount of information produced as well as consumed in the world is constantly expanding. Users anticipate their applications to cope with such an expansion, demanding the support of novel and heterogeneous data sources and collections. At the same time, due to the advanced connectivity offered by the expansion of the Internet, users need to use the same information in different contexts, reusing and refining data across applications. In this thesis, we deal with the common challenge underlying the above data expansion, integration and interoperability needs, which is how to automate the inclusion of new data in existing service provisions. To avoid invasive, time-consuming and expensive source-code extensions that frequently break applications, we propose the virtualization of information spaces, introducing the notion of virtual objects. Our proposal seeks to shift the barriers raised by heterogeneous data representations, formats and protocols by ascertaining the following assumption. Should we manage to isolate the "substance" of data from any application-specific "materialized forms", we can automate the process of detaching the data from one context and attaching the data to another, in a multitude of operational environments. Virtual objects are logically placed between the business-logic and the data-sources of applications to offer a common, reusable and composable runtime interpretation of data that transcends the application-context boundaries. On one hand, proposed virtual objects allow applications to adapt to new information, without changing their business-logic implementation. On the other, virtual objects can inter-connect data spaces to facilitate the integration and interoperability of information. We expect that data-centric applications will benefit from our approach, while our experimentation shows that virtual objects impose minimal overheads even when used atop heterogeneous sources.

Keywords: Virtual Objects, Information Expansion, Data Integration, Interoperability, Middleware, Domain-specific Languages

* Dissertation Advisor: Alex Delis, Professor

1 Introduction

The concept of *virtualization* originates from operating systems [1], where a software abstraction layer is introduced—the virtual machine—to behave as a computer system's hardware—a real machine. With the emergence of cloud computing and the "software as a service" paradigm, virtualization has offered flexibility, automation and ease of use, reducing OS installation, deployment and maintenance costs [2]. In addition, programming platforms such as Java or .NET have built upon virtual machines to offer features like architecture neutrality, garbage collection and network code loading, ultimately advancing developer productivity.

In this thesis, we use virtualization as a means to *enable the use of information spaces to transcend the application-context boundaries*, allowing users to create, share, reuse and refine information across applications. To this end, we work to automate the process of detaching an information space from one context and re-attaching the space to another in various operational environments. Should we view such data-detach and data-attach as logical operations, at a high level of abstraction, these operations play a significant role in the information space life-cycle, since information spaces:

a. *constantly expand*: automating the ability to attach new information spaces to existing application contexts will allow for the cost-effective introduction of novel types of content in applications and, thus, foster the expansion of information.
b. *outlive applications*: automating the ability to detach information spaces from their applications will simplify data migration and also reduce the data lock-in by legacy technology.
c. *should interoperate*: automating the ability to detach/attach information spaces from/to application contexts will clearly simplify interoperability.

In order to separate the information space from the application context, and offer reusable and composable data detach/attach operations, we propose *the virtualization of the information space*. The unique characteristic of our approach is that we use proposed *virtual objects* as archetypes of data items, offering a platonic view of data [3]; a virtual object aims to identify the "substance" of a data item, independently of any application-specific "materialized forms" such as storage-specific data layouts, protocol-based data representations or application-oriented data views. Our objective is to offer a software abstraction layer—a virtual machine for data objects—which is embedded in applications in order to enable the use of information to transcend the application-context boundaries.

1.1 Problem Definition

In working with virtual objects, our contribution is that we simplify and automate a wide variety of real-world data-expansion, migration and interoperability issues. Specifically, we manage to answer the common challenge underlying these

issues, which is *how to automate the inclusion of new data in existing service provisions*. Data-expansion calls for augmenting the information space with new data sources and collections, coping with the pressure to rapidly adapt to new information. In data-migration, the difference is that the new data source has to be migrated before participating in the existing service provision. In data-interoperability, the difference is that applications use a data-exchange protocol to act as the data sources of each other. Yet, in all cases, similar, if not identical, steps need to be performed, including:

Step 1: revisit data access actors to support newly encountered data sources or collections. For example, to support a new data source, an application may need to introduce a new machinery to access the new data, dealing with network connections, database queries, XML parsing, web services and other protocols. Similar revisions may need to be performed in order to support a novel collection type.

Step 2: introduce new business-logic actors to stage the new data. For example, a Java application has to introduce new Java classes and objects to stage the newly encountered data items at runtime.

Step 3: extend the current service provision to deal with the new data. For example, the application has to include the new business-logic actors in its existing implementation of services.

By virtualizing the information space, we work towards automating these steps in any operational environments. Viewed from different perspectives, the above steps raise various multi-disciplinary data integration, data quality and software evolution/adaptation issues [4–7]. Clearly, developers can handle these cases by code re-engineering. However, data expansion or integration requirements cannot always be predicted in detail during the initial application design and development phases. Consequently, the support of a new type of content may break the application, imposing drastic, invasive and expensive source code changes to all service provision actors. Developers may follow an ad-hoc approach to revisit service provision implementation and they can ultimately succeed in including this new content. However, yet another new requirement for supporting additional types of content may render this approach problematic, breaking service provision actors yet again.

The crucial need here is not to predict the future, but rather to achieve the return of investment on existing services; indeed, the challenge is to enable the existing service provision to operate atop constantly expanding, inter-connected and heterogeneous information spaces. Thus, a better approach is to base the application on a flexible framework that can isolate the application logic from the type of context and add indirection between the business-logic and the information space. Although adding indirection is a simple idea, designing a general and flexible framework for data expansion and integration is anything but simple. The framework needs to: a) isolate the structure of data, i.e., how the logical organization of data (e.g., the tuples of a database, or the elements of XML documents) map to the application's expectations; b) adapt the physical access to

data (e.g., provide network or database connections to objects in a way transparent to the application); c) abstract the object presentation, i.e., smoothly integrate the display of new kinds of objects in the application user interface; d) abstract the object manipulation, i.e., allow new object modification in a uniform way; and e) perform these tasks conveniently and efficiently, in particular without imposing significant runtime overhead over an inflexible, hard-coded implementation of the same features.

1.2 Contribution

Our proposal meets the above challenging requirements by virtualizing the information space. We use the term *information space* to refer to the dataset being managed by an application. Given that different applications develop different data manifestations and issue different mechanisms for their management, such application-specific characteristics designate the *application-specific information context*, or simply the *application-context*. The proposed virtual object environment decouples the information space from the application-context, achieving the multi-dimensional separation [8] of the following concerns:

1. *data-access/storage*: the storage representation(s) of data,
2. *data-synthesis*: the usage and composition of data,
3. *data-conceptualization*: the logical structure and modeling of data,
4. *data-discovery*: the data searching/indexing functionality.

These four concerns reflect the essential dimensions that couple the data to the application. By loosening these couplings, we manage to treat the four dimensions that compose the data/application interactions in in an orthogonal manner. Such a separation virtualizes the information space, automating the three data expansion/integration steps, by offering a *common interpretation of heterogeneous and diverse data* that transcends the application-context boundaries. Our proposal is well-aligned with the long-term objective of data independence [9, 10] and contributes to the open challenge of interpreting and synthesizing heterogeneous data in an automated manner [11, 12]. To the author's knowledge, the proposed virtual object approach is the first to deal with data expansion, interoperability and migration in a unified and integrated way. The majority of our work with virtual objects appears in [13–17] and [18].

Our implementation of virtual objects, called *DOLAR (Data Object Language And Runtime)*, is realized in Java and consists of a virtual object domain-specific embeddable language (DSEL) [19–21] and the respective runtime environment. We have used DOLAR in several real-world applications and operational environments, allowing us to effectively answer various data expansion, integration and interoperability needs. For example, DOLAR has allowed us to deal with data-expansion in Pergamos, the Univ. of Athens digital library (http://pergamos.lib.uoa.gr (http://pergamos.lib.uoa.gr)). Pergamos is the largest academic digital library in Greece, hosting about 300,000 items and exceeding 1 TB of space, while it has been in production use for more than five years. In

the thesis, we present how the use of DOLAR has helped us answer the need to gradually (a) use Pergamos to develop a variety of collections originating from independent digitization projects at the University, (b) add existing University collections in Pergamos, including digitized books, Domino-based dissertations, etc. The use of virtual objects has achieved the return of investment on Pergamos services, allowing the business-logic to adapt to the gradual inclusion of new data sources and collections without modifications. In general, embedding virtual objects in applications dissociates the application-logic from the data-inherent idiosyncrasies, enabling applications to extend their "low-level" information space options without modifying their "high-level" business-logic services. In terms of cross-context DOLAR usage, we have used virtual objects in a real-world interoperability scenario originating from John S. Latsis Public Benefit Foundation (http://www.latsis-foundation.org). We show how the virtual space has allowed the Foundation's digital archive and its collections to transcend the application boundaries, presenting the effectiveness, automation and reuse offered when dealing with data migration and interoperability issues. Finally, our experimental evaluation, carried out in both synthetic and production environments, shows that DOLAR-imposed operational overheads are minimal; DOLAR-enabled applications scale as well as the underlying datastore(s), even when used atop a variety of heterogeneous SQL, XML and Web data sources.

2 Related Work

Interoperability is a strong and diachronic requirement of data-intensive systems. Syntactic interoperability refers to the ability of systems to communicate and exchange data, while semantic interoperability refers to the ability of systems to accurately and meaningfully interpret and use the data in an automated manner [11, 12]. XML is a flexible markup language, offering an extensible representation to store and exchange data. To this effect, XML has been used to deliver a plethora of protocols for achieving syntactic interoperability. The Semantic Web [22] uses RDF and ontologies [23] to build upon XML to offer semantic interoperability. However, as put in [24], it may be that there are many semantic webs. Indeed, various middleware [25, 26] and service-oriented architectures [27] are used to consolidate heterogeneous APIs and services to make them interoperate in various operational environments. Although interoperability involves various cultural, social and legal issues [28], the key challenge is to achieve automation, enabling information that originates in one context to be used in another in ways that are as highly automated as possible [29].

This is the exact goal of our approach, decoupling the information space from the application-context to permit the use of data to transcend the application-context boundaries. Our virtual space proposal builds upon the separation of concerns to loosen the data/application couplings and automate the cross-context usage of data. The key here is that the proposed virtual space offers a uniform means to deal with the three data expansion/integration steps, which are relevant to all kinds of applications and all types of data. The virtual object space

can be used in conjunction with any application architecture and in the thesis we demonstrate the use of DOLAR in two production cases, where virtual objects are used in an MVC [30] and a REST[31] architecture.

Our approach to realize the virtual space as a DSEL aims to achieve an additional level of separation, that is, to decouple the data from the application-context in terms of the programming-language used. Our long-term goal with virtual objects is to offer an embeddable, language-independent, data-centric runtime environment—*a virtual machine for data objects*. Although our current DOLAR implementation is realized in Java, the key elements of the virtual space—including virtual objects and prototypes, datastore drivers and connectors, composition schemes and virtual object inheritance—can be clearly implemented in any general-purpose programming-language. Moreover, the prototype format feature of our DSEL builds upon the introspection of DOLAR prototypes to bridge virtual objects with multiple syntactic notations, including JSON, XML or our custom DOLAR syntax. Different applications can use different notations for storing and exchanging their prototypes, yet, all notations are parsed to generate identical prototypes in terms of the DOLAR language. This way, we manage to treat the data exchange and definition syntax as a pluggable component of the virtual space and not as a hard-coded option. In general, the DOLAR DSEL combines the flexibility and versatility of XML with the programmability and ease of use of scripting languages, in an effort to offer a foundation for issuing multiple data-definition utilities; for example, we also present *DOPs Creator*, a GUI tool for defining prototypes.

Finally, XML ontologies are widely used to model and conceptualize information in various contexts [32]. The main difference between ontologies and our approach—which stands for any comparison between ontologies and OO systems—is that ontologies use inference as the primary compositional mechanism, while in DOLAR we use virtual object instantiation and inheritance.

3 Overview of the Virtual Object Space

A virtual object offers a runtime manifestation of data which separates (a) how the data are being accessed and stored, (b) how the data are logically arranged at runtime and (c) how the data are synthesized and composed. Figure 1 depicts a virtual object, comprising (from left to right):

a. an application-specific physical/storage representation of data. The virtual object uses our *datastore driver (DOStore)* mechanism to capture the physical data access/storage options of applications.
b. an application-neutral logical structure of data. The virtual object conforms to a conceptualization expressed in terms of a *structural prototype* definition, offering a uniform, application-neutral representation of the logical arrangements of data at runtime.
c. an application-specific behavior of data. The virtual object adheres to a *behavioral prototype*, yielding application-specific sub-objects which are compatible with the application's service provision expectations.

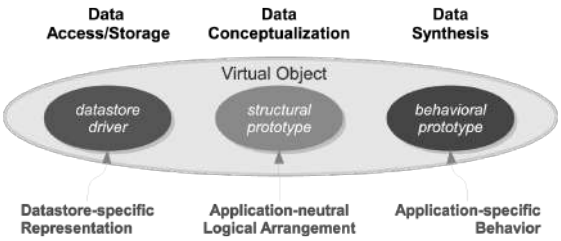


Fig. 1. A virtual object comprising a datastore driver, a structural and a behavioral prototype

By separating the left, data storage concern from the middle, data conceptualization concern, we manage to virtualize the datastores and treat heterogeneous storage artifacts as serializations of virtual objects. A virtual object adheres to the logical structure defined by its structural prototype consisting of *fields*, used for holding attributes, metadata or similar name/value pairs; *relation contexts*, used for holding the relationships among objects; and *stream handles*, used as pointers to files and similar document-based content. In OO terms, these elements designate the internal state of virtual objects; the structural prototypes provide the logical arrangements of data, offering the "data object classes", and virtual objects provide the runtime manifestation of data, offering the "data object instances". During instantiation, virtual objects use the *datastore driver* mechanism to establish a two-way link to the underlying storage artifacts. The driver connects the virtual field, relation-context and stream-handle runtime structures to any respective storage-specific structures employed by the datastore(s) used beneath. The datastore driver mechanism designates the data-access API of the virtual space, where the datastore drivers play the role of the data-access actors of Step 1. The virtual space uses the driver mechanism in a transparent to the developer fashion, offering an effective virtual object load/store metaphor which virtualizes the datastores. For example, developers can both load and store heterogeneous virtual objects using literally a single line of code. The developer is provided with the abstraction that heterogeneous storage artifacts are serializations of virtual objects, dealing with runtime artifacts that can access and also update any datastores with identical effectiveness and ease of use. Behind the scenes, the virtual space automatically deals with common data fetching tasks including (a) staging the data in virtual object runtime structures, (b) synchronizing the access to these structures and, finally, (c) flushing such structures to underlying datastores as needed.

By separating the middle, data-conceptualization concern from the right, data-synthesis concern, we separate the data-inherent from application-inherent behavior as follows:

– *data-inherent behavior* depends on the logical arrangements of the data in question; for example, the thumbnail of a "book" item may originate from

a digitized image of the book's first page, while the thumbnail of a "photo" item may originate from the digitized image of the photo itself.
- *application-inherent behavior* depends on the service provision requirements of the application at hand, regardless of the data items supported; for example, an application will either provide a thumbnail display for all of its items or it will not support such a display at all.

As far as the business-logic is concerned, the acquisition of a thumbnail from a photo, a book, or any other item should be transparent. The interpretation of data-inherent behavior (i.e. how to acquire the thumbnail) depends on the structure of the data item at hand. The implementation of application-inherent behavior (i.e. what to do with the thumbnail) depends on the service provision at hand. We use the notion of *composition schemes* to achieve this separation, offering an amalgamation of (a) a method construct of OO languages and (b) a projection operation of query languages. There is a direct analogy between the methods of "code objects" and the composition schemes of virtual objects: the methods/schemes available on an object define the messages the object can respond to. Yet, in contrast to "code objects", a virtual object *does not contain executable code* but responds to messages by transforming its internal state. A behavioral prototype defines a set of composition schemes, designating the runtime interface between virtual objects and application artifacts such as components and modules. The analogy, in a general-purpose OO language, is that schemes correspond to methods and behavioral prototypes correspond to interfaces. Schemes are realized and behave at runtime as sub-objects, transforming the structure of virtual objects to automatically match the application's expectations. Composition schemes are are a unique feature of our approach, providing the key for enabling virtual objects to expose different behavior when used across applications.

The virtual object environment is logically placed between the business-logic and the data sources to offer:

1. *a common representation of heterogeneous data*. For example, heterogeneous storage artifacts, such as SQL "book" tuples and XML "book" documents, will adhere to a uniform "book" virtual object representation at runtime. This is achieved with the help of the virtual object load/store metaphor, as realized by virtual object prototypes and datastore drivers described before. This metaphor enables us to treat heterogeneous storage artifacts as native objects of the virtual space, simplifying Steps 1 and 2.
2. *a uniform composition of semantically different data*. For example, diversely structured data, such as "book" and "photo" items, regardless of their storage representations, will conform to uniform manipulation at runtime. Here, we build upon the common representation offered by the virtual space to manage diversely structured virtual objects through a uniform programming interface. This is critical, as we seek objects that can be defined by their responses to messages and not by their internal representation [33]. With the use of our composition schemes and behavioral prototypes, the service provision logic

can catch up with the gradual addition of new content without source code modifications, automating Step 3.

3. *a set of reusable and composable data attach/detach mechanisms.* For example, applications can build upon the virtual objects of each other, regardless of their network location, logical structure or storage representation, to share, reuse and refine their data spaces. The data attach/detach mechanisms of DOLAR include *virtual object inheritance* and the *virtual space connector* and *prototype format* mechanisms. On one hand, prototype formats and virtual space connectors offer a pluggable and disciplined means to (a) exchange prototypes and (b) inter-connect virtual spaces, allowing DOLAR applications to instantiate the virtual objects of each other. On the other hand, virtual object inheritance allows for the reuse and refinement of virtual object definitions, automating the inclusion of new data in existing service provisions.

Building upon such virtual object structural and behavioral prototypes, we offer a mixin-based [34–36] multiple inheritance mechanism. With virtual object inheritance, we support the subclassing, subtyping and specialization features offered by inheritance in general, yet, we use these features to automate the cross-context usages of data. For example, virtual objects can be polymorphic in the OO sense, yet, the virtualization offered by the proposed environment allows virtual objects to effectively support different polymorphisms in different contexts. Virtual object inheritance provides the most powerful data attach/detach mechanism of our proposal, as it enables developers to reuse and extend virtual objects for attaching/detaching information among different contexts.

4 Conclusions

Our proposal offers a novel virtual object language and runtime that transcends the data, knowledge and software engineering boundaries in order for information spaces to transcend the storage, programming-language and application boundaries. We view our virtual object approach as an effort to offer a virtual machine for data objects, an embeddable environment that permits applications to develop a common, reusable and composable interpretation of data that transcends the application-context boundaries. The virtual space can be embedded in applications; it can reused across contexts; it can be refined to effectively match different service provision expectations. The virtual view of data employed in our approach allows us to deal with the common challenge underlying data expansion, integration and migration, that is, the adaption of existing service provisions to new information. As the amount of information undergoes constant growth globally, the above adaptation requirement becomes more and more dominant, as it involves all types of applications and all kinds of data. The proposed virtual objects meet this requirement by allowing developers to expand applications to support novel information without breaking their existing implementations.

References

1. R.P. Goldberg. Survey of virtual machine research. *IEEE Computer Magazine*, pages 34–45, June 1974.
2. M. Rosenblum. The reincarnation of virtual machines. *Queue*, 2(5):34–40, 2004.
3. D. Ross. *Plato's Theory of Ideas*. Oxford University Press, 1951.
4. M. Lenzerini. Data integration: a theoretical perspective. In *PODS '02: Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 233–246, New York, NY, USA, 2002. ACM.
5. K. Chen, H. Chen, N. Conway, H. Dolan, and J. M. Hellerstein T. S. Parikh. Improving data quality with dynamic forms. In *ICTD'09: Proceedings of the 3rd international conference on Information and communication technologies and development*, pages 487–487, Piscataway, NJ, USA, 2009. IEEE Press.
6. I. Lukovic, P. Mogin, J. Pavicevic, and S. Ristic. An approach to developing complex database schemas using form types. *Softw., Pract. Exper.*, 37(15):1621–1656, 2007.
7. D. Parsons, A. Rashid, A. Telea, and A. Speck. An architectural pattern for designing component-based application frameworks. *Softw. Pract. Exper.*, 36(2):157–190, 2006.
8. P. Tarr, H. Ossher, W. Harrison, and S.M. Sutton. N degrees of separation: Multi-dimensional separation of concerns. In *Proc. of the 21st Int. Conf. on Software Engineering (ICSE)*, pages 107–119, 1999.
9. C. J. Date and P. Hopewell. File definition and logical data independence. In *Proceedings of the 1971 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control*, SIGFIDET '71, pages 117–138, San Diego, California, 1971. ACM.
10. J.M Hellerstein. Toward network data independence. *SIGMOD Rec.*, 32(3):34–40, September 2003.
11. S. Heiler. Semantic interoperability. *ACM Comput. Surv.*, 27(2):271–273, 1995.
12. A.M. Ouksel and A. Sheth. Semantic interoperability in global information systems. *SIGMOD Rec.*, 28(1):5–12, 1999.
13. K. Saidis, G. Pyrounakis, and M. Nikolaidou. On the effective manipulation of digital objects: A prototype-based instantiation approach. In *Proceedings of the 9th European Conference on Digital Libraries*, pages 26–37, Vienna, Austria, 2005.
14. K. Saidis, G. Pyrounakis, M. Nikolaidou, and A. Delis. Digital object prototypes: An effective realization of digital object types. In *Proceedings of the 10th European Conference on Digital Libraries*, Alicante, Spain, September 2006.
15. K. Saidis and A. Delis. Towards a Unified Runtime Model for Managing Networked Classes of Digital Objects. In *2nd DELOS Workshop on Foundations of Digital Libraries, In conjunction with the 11th European Conference on Digital Libraries*, Budapest, Hungary, 2007.
16. K. Saidis and A. Delis. Type-consistent Digital Objects. *D-Lib Magazine*, 13(5/6), May/June 2007. [doi:10.1045/may2007-saidis].
17. K. Saidis and A. Delis. Integrating multi-dimensional information spaces. In *2nd Workshop on Very Large Digital Libraries, In conjunction with the 13th European Conference on Digital Libraries*, Corfu, Greece, 2009.
18. K. Saidis, Y. Smaragdakis, and A. Delis. Dolar: virtualizing heterogeneous information spaces to support their expansion. *Software: Practice and Experience*, Accepted for publication, 2010, available online at http://dx.doi.org/10.1002/spe.1050.

19. P. Hudak. Building domain-specific embedded languages. *ACM Computing Surveys*, 28(4es), 1996.
20. D. S. Wile. Supporting the DSL Spectrum. *Journal of Computing and Information Technology*, CIT 9(4):263–287, 2001.
21. M. Mernik, J. Heering, and A.M. Sloane. When and how to develop domain-specific languages. *ACM Computing Surveys*, 37(4):316–344, 2005.
22. N. Shadbolt, T. Berners-Lee, and W. Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006.
23. S. Decker, S. Melnik, F. van Harmelen, D. Fensel, M. C. A. Klein, J. Broekstra, M. Erdmann, and I. Horrocks. The semantic web: The roles of xml and rdf. *IEEE Internet Computing*, 4:63–74, 2000.
24. C. Marshall and F. Shipman. Which semantic web? In *HYPERTEXT '03: Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, pages 57–66, New York, NY, USA, 2003. ACM.
25. P. A. Bernstein. Middleware: a model for distributed system services. *Commun. ACM*, 39:86–98, February 1996.
26. F. Kon, F. Costa, G. Blair, and R. H. Campbell. The case for reflective middleware. *Commun. ACM*, 45:33–38, June 2002.
27. T. Erl. *Service-Oriented Architecture: Concepts, Technology, and Design*. Prentice Hall, 2005.
28. P. Miller. Interoperability. What is it and Why should I want it? *Ariadne, Issue 24*, June 2000. http://www.ariadne.ac.uk/issue24/interoperability/intro.html.
29. IDF. The DOI Handbook, The International DOI Foundation, 2006. Edition 4.4.1, October 2006, [doi:10.1000/182].
30. E. Gamma, R.Helm, R.Johnson, and J.Vlissides. *Design Patterns Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1997.
31. R. T. Fielding and R. N. Taylor. Principled design of the modern web architecture. In *ICSE '00: Proceedings of the 22nd international conference on Software engineering*, pages 407–416, New York, NY, USA, 2000. ACM.
32. B. Chandrasekaran, John R. Josephson, and V. Richard Benjamins. What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1):20–26, 1999.
33. H. Lieberman. The continuing quest for abstraction. In *Proceedings of the 20th European Conference on Object Oriented Programming (ECOOP)*, pages 192–197, 2006. doi: 10.1007/11785477_12.
34. G. Bracha and W. R. Cook. Mixin-based Inheritance. In *OOPSLA / ECOOP*, pages 303–311, 1990.
35. Y. Smaragdakis and D. Batory. Implementing layered designs with mixin layers. In *Proceeding of the European Conference on Object Oriented Programming (ECOOP)*, 1998.
36. N. Scharli, S. Ducasse, O. Nierstrasz, and A. Black. Traits: Composable units of behaviour. In *ECOOP 2003 – Object-Oriented Programming*, volume 2743 of *Lecture Notes in Computer Science*, pages 327–339. Springer Berlin / Heidelberg, 2003.

Quality of Service Provision for IP Traffic over Wireless Local Area Networks

Dimitris Skyrianoglou[1]

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
dimiski@di.uoa.gr

Abstract. This PhD dissertation deals with the provision of guaranteed quality of service (QoS) to the users of a Wireless LAN (WLAN) and the interworking between WLANs and the 3rd generation and IP networks. The work is divided into three main parts:
i) Study and development of Wireless Adaptation Layer (WAL), a shim, transparent –both for the IP layer and the underlying WLAN- layer that supports the interworking of WLANs with IP networks and provides guaranteed QoS over WLANs by utilizing the QoS mechanisms of WAL..
ii) Study of traffic scheduling algorithms for WLANs that are based on IEEE 80211e protocol, the extension of legacy IEEE 802.11 protocol for supporting guaranteed QoS over 802.11 WLANs. More specifically a new traffic scheduling algorithm for 802.11e called ARROW (Adaptive Resource Reservation Over Wireless) and an extension of ARROW called ARROW (P-ARROW) was developed and evaluated.
iii) Study of interworking between WLANs and UMTS for the provision of guaranteed QoS for the mobile users that perform a handover from one network to the other. The focus was on how the QoS mechanisms of UMTS and WLANs can interwork and combine so as to offer guaranteed QoS service to the users that perform a handover.

Keywords: Quality of Service (QoS), DiffServ, Wireless Adaptation Layer (WAL), IEEE 802.11e, Traffic Scheduling Algorithms, ARROW Scheduler, UMTS/WLAN Interworking, Seamless Handover

1 Introduction

The rapid development and the high transmission rates attained by the Wireless Local Area Networks (Wireless LANs - WLANs) have established them as one of the most attractive choices for supporting alternative access to large 3rd generation networks (3G) like UMTS or metropolitan IP networks. The installation of WLANs in places with a dense mobile user population (i.e. hot-spots like malls, airports, hospitals etc.) relieves the traffic load towards the metropolitan networks while, at the same time, achieves an improved level of quality of service for the mobile users.

[1] Dissertation Advisor: Lazaros Merakos, Professor

The work in hand deals with the provision of guaranteed quality of service (QoS) to the users of a WLAN and the interworking between WLANs and the 3rd generation and IP networks. The provision of quality of service to WLAN users at a level at least equal to that offered by the metropolitan network is deemed as especially important since the objective is to offer the mobile users a uniform level of quality of service regardless of their current location.

In this respect this work proposes the introduction of a new shim-layer called *Wireless Adaptation Layer (WAL)* that lies between the IP and the underlying wireless LAN DLC layer and aims at providing or complementing the QoS support for the underlying WLAN platform [1]-[3]. Further to this, the work delved into the QoS support mechanism of IEEE 802.11e WLAN protocol and proposed a novel traffic scheduling algorithm named *ARROW (Adaptive Resource reservation Over Wireless)* together with an extension of ARROW called *P-ARROW (Prioritized-ARROW)* [4]-[8]. Finally the work examined the interworking of WLANs with 3G networks like UMTS focusing again on the provision of QoS and proposing an architecture for supporting seamless handover for voice and video streams from one platform to the other [9]-[15].

1.1 Wireless Adaptation layer (WAL)

Several solutions are available in the literature, coping with limitations of the wireless links. Most of these solutions propose enhancements at the Transport or Application layers, while others focus on the Link Layer trying to transparently improve higher layers performance and thus avoid modifications. A number of these solutions fall into the category of Performance Enhancing Proxies (PEPs) that are defined as elements used to improve the performance of Internet Protocols on network paths where native performance suffers due to characteristics of a link or subnetwork path.

The approach proposed in this work is in line with the idea of PEPs but also tries to expand and generalize it. More specifically, it is based on the introduction of an intermediate layer called Wireless Adaptation Layer (WAL) between the IP and the Link Layer. WAL incorporates a set of functional modules, viewed as generalized PEPs, that can be dynamically combined and adapted to the special characteristics of the wireless link and the transport protocol.

WAL architecture is shown in Fig 1. A novel and key feature of the WAL is that it is an abstraction used for service provisioning at the link layer [1]-[3]. Each IP packet is classified by WAL into classes and associations. A WAL class defines the service offered to a particular set of IP packets and corresponds to a particular sequence of WAL modules that provide such a service. A WAL association identifies a stream of IP packets classified for the same WAL class and destined to or originated from a specific mobile terminal (MT). In other words, a WAL association corresponds to a particular type of service offered to a particular MT. In this way, we can differentiate the operation of WAL on a per-user basis. In addition, services for particular users can be customized to meet their specific QoS requirements and to implement a differentiated-charging policy.

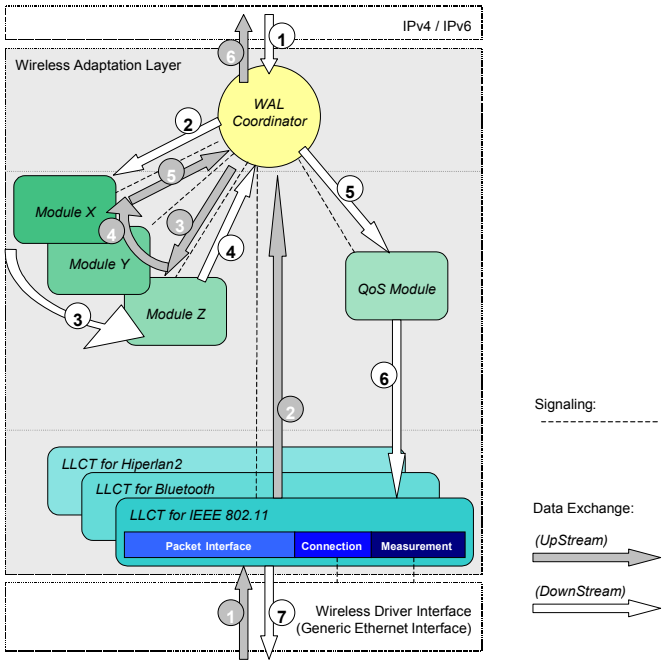


Fig. 1. WAL Architecture

The WAL Coordinator shown in Fig. 1 can be viewed as the central "intelligence" of the WAL. Both downstream (from IP layer) and upstream (to IP layer) traffic passes through the WAL Coordinator before being processed by other modules.

The QoS module (shown in Fig. 1) provides flow isolation and fairness guarantees through traffic shaping and scheduling. On the other hand, modules X/Y/Z comprise a pool of functional modules, aiming to improve performance in a number of ways. The modules that have been identified so far are: ARQ module, FEC module, Fragmentation module, IP Header Compression module, and SNOOP module.

Finally, in order to interface with a number of wireless drivers of different wireless technologies (such as IEEE 802.11, Bluetooth, HiperLAN/2, etc.), one Logical Link Control Translator (LLCT) module for each different wireless technology has been introduced. The main functions of this module manage the connection status with the wireless driver, and ensure the stream conversions toward the wireless driver.

For the classification of the IP packets to WAL classes a service differentiation is needed. Service differentiation in WAL is based on the DiffServ architecture. In this respect, the wireless access system can be viewed as a DiffServ domain with the Access Point acting as the DiffServ boundary node, interconnecting the wireless access system with the core network or other DiffServ domains.

1.2 Traffic Scheduling in IEEE 802.11e

The IEEE 802.11 standard is considered today the dominant technology for wireless local area networks (WLANs). Besides great research interest, 802.11 has enjoyed widespread market adoption in the last few years, mainly due to low-price equipment combined with high bandwidth availability. Recent improvements in the physical (PHY) layer provide transmission speeds up to hundreds of Mbps per cell, facilitating the use of broadband applications. However, one of the main weaknesses of the original 802.11, towards efficient support of multimedia traffic, is the lack of enhanced Quality of Service (QoS) provision in the Medium Access Control (MAC) layer.

In order to eliminate these weaknesses and respond to business requirements for multimedia over WLANs, IEEE is currently working on a set of QoS-oriented specification amendments, referred to as IEEE 802.11e, that enhance the existing MAC protocol and facilitate the multimedia QoS provision. In IEEE 802.11e, the QoS mechanism is controlled by the Hybrid Coordinator (HC), an entity that implements the so-called Hybrid Coordination Function (HCF). The HC is typically located in an Access Point (AP) and utilizes a combination of a contention-based scheme, referred to as Enhanced Distributed Coordination Access (EDCA), and a polling-based scheme, referred to as HCF Controlled Channel Access (HCCA), to provide QoS-enhanced access to the wireless medium. EDCA provides differentiated QoS services by introducing classification and prioritization among the different kinds of traffic, while HCCA provides parameterized QoS services to Stations (QSTAs) based on their traffic specifications and QoS requirements. To perform this operation, the HC has to incorporate a scheduling algorithm that decides on how the available radio resources are allocated to the polled QSTAs. This algorithm, usually referred to as the Traffic Scheduler, is one of the main research areas in 802.11e, as its operation can significantly affect the overall system performance [4]. Traffic Schedulers allocates resources to the QSTAs in the form of Transmission Opportunities (TXOPs). A TXOP is an interval of time when a QSTA obtains permission to transmit onto the shared wireless channel.

In the open technical literature, only a limited number of 802.11e traffic schedulers have been proposed so far and this work partially aims at filling this gap. The draft amendment of IEEE 802.11e includes an example scheduling algorithm, referred to as the Simple Scheduler, to provide a reference for future, more sophisticated solutions. The idea of this algorithm is to schedule fixed batches of TXOPs at constant time intervals. Each batch contains one fixed length TXOP per QSTA, based on the mean data rates as declared in the respective Traffic Specifications (TSPECs). With this discipline the Simple Scheduler respects the mean data rates of all TSs and performs well when the incoming traffic load does not deviate from its mean declared value (e.g., constant bit rate traffic). On the other hand, its performance deteriorates significantly when it comes to bursty traffic, as it has no means to adjust TXOP assignments to traffic changes. Identifying the weaknesses of the Simple Scheduler, SETT-EDD (Scheduling based on Estimated Transmission Times - Earliest Due Date) scheduler provides improved flexibility by allowing the HC to poll each QSTA at variable intervals, assigning variable length TXOPs. With SETT-EDD TXOP assignments are based on earliest deadlines, to reduce transmission delay and packet

losses due to expiration. SETT-EDD is a flexible and dynamic scheduler, but it lacks an efficient mechanism for calculating the exact required TXOP duration for each QSTA transmission. Each TXOP duration is estimated based on the mean data rate of each TS and the time interval between two successive transmissions.

In order to overcome the disadvantages of Simple and SETT-EDD schedulers this work proposes a new scheduling algorithm, referred to as *Adaptive Resource Reservation Over WLANs (ARROW)* [4], [6], [7], that adapts TXOP durations based on the backlogged traffic reports issued by QSTAs. The novel characteristic of ARROW is that it exploits the Queue Size (QS) field, introduced by 802.11e as part of the new QoS Data frames, not supported by legacy 802.11 systems. The QS field can be used by the QSTAs to indicate the amount of buffered traffic for their TSs, i.e., their transmission requirements. Furthermore, in order to take advantage of the periodic nature of CBR streams, a CBR-enhancement of ARROW was also developed.

Simulation results show that ARROW achieves much more efficient use of the available bandwidth, compared to Simple and SETT-EDD, leading to better channel utilization and higher throughput. The increased transmission overhead percentage of the proposed scheduler turned to be not a significant performance issue. Finally, it is important to note that ARROW does not mandate any standards changes. It could be readily deployed and implemented in practice, provided that STAs populate the QS field as defined in the 802.11e standard.

An important extension of ARROW is *P-ARROW (Prioritized ARROW)* [5]. The main enhancement of P-ARROW compared to ARROW is its ability to efficiently handle different traffic classes. The novel characteristic of P-ARROW is the introduction of Priority Factor a, and the use of traffic priorities based on delay constraints. Performance results extracted from simulation models, show that P-ARROW is very efficient in supporting the desired level of service differentiation and prioritization among different traffic classes.

1.3 UMTS/WLAN Interworking

As the Internet technologies evolve, more sophisticated and Quality of Service (QoS) demanding multimedia services are being requested by the users. The Internet Protocol (IP), together with its QoS enhancement frameworks (namely the Integrated Services - IntServ - and the Differentiated Services - DiffServ), is currently the main transport technology for supporting all these new services and in this respect the motto “Everything over IP and IP over everything” has become the trend of the day. On the other hand, both UMTS and Wireless LANs (WLANs) are already commercially available and become increasingly popular. The number of mobile users is growing rapidly and so does the demand for wireless access to the Internet services, imposing the need for a unified QoS support framework in both UMTS and WLANs.

Despite the initial impression, expressed by several network technology vendors, that UMTS and WLANs will be competing technologies it appears that they can be combined and complement each other in an effective way. The approach followed in this work is that both UMTS and WLANs can act as access systems to one common

IP core network, efficiently covering both wide areas and hot-spots. One of the main requirements of this system is a unified QoS support for IP traffic. As RSVP is considered the dominant signaling protocol of IP traffic, the discussion focuses on the adoption of RSVP messages and parameters by UMTS or WLAN QoS mechanisms [12]-[15].

Further to this, as the next-generation networks (NGN) are expected to support a wide variety of service types, especially broadband multimedia services, including video conference, streaming, and advanced telephony services, a major objective is how these services should operate seamlessly across all diverse access systems (e.g. WLAN, UMTS, fixed broadband, WiMAX, cable, etc). This seamless operation presents several challenges especially when the different access systems are loosely coupled and therefore lack the tight integration we experience in GSM/UMTS radio environments for instance. To address this issue for the case of UMTS/WLAN interowrking this work proposes a specific architecture for the support of seamless voice and video handover between the two platforms [9]-[11]. The basic idea of the proposed architecture is that a new internal entity of UMTS calles Seamless Handover Control Function, located at the IMS (Internet-Multimedia System) will act as an anchor-point hiding user mobility from the external IP network. Both UMTS and WLAN are also equipped with appropriate entities that take care of interworking procedures such as re-routing of traffic and authentication of the roaming users. The proposed architecture for the case of seamless voice handover is depicted in Fig. 2.

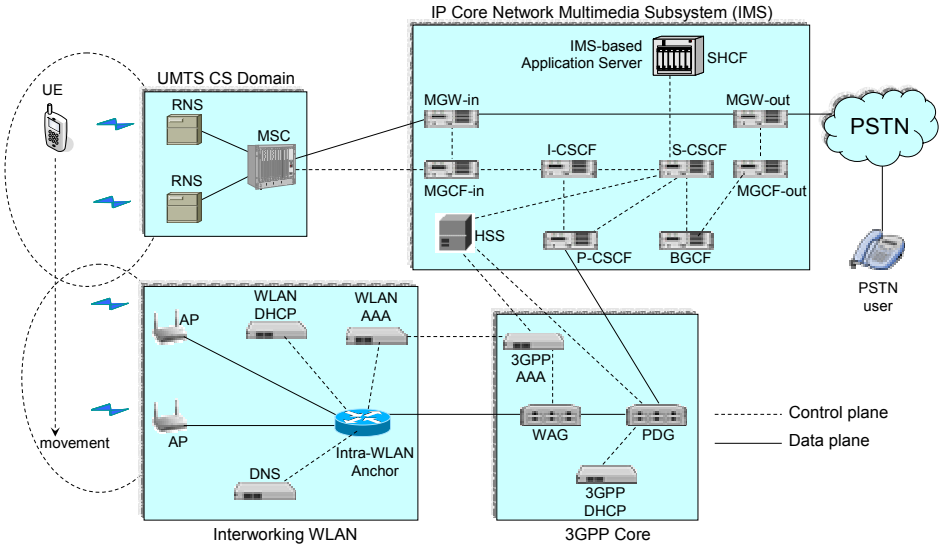


Fig. 2. IMS-based architecture for enabling seamless voice handover across UMTS and WLAN.

Simulation results indicate that WLAN can accommodate a limited number of UMTS roamers (i.e. users that perform a handover from UMTS to WLAN). This number depends on the bandwidth allocated for these users, their QoS requirements and also on the QoS support mechanism of WLAN [9]-[11].

2 The ARROW Scheduler

In IEEE 802.11e the traffic scheduler has to decide on the next TXOP assignment taking into account traffic characteristics and QoS requirements expressed through TSPEC parameters. As already mentioned, TXOP assignments are performed per QSTA, while TSPECs are defined per TS. Therefore, for each $QSTA_i$ having n_i active TSs (where i is the index of the QSTA), the traffic scheduler has to utilize some aggregate parameters, derived from the individual TSPECs, which are calculated as follows:

Minimum TXOP duration (*mTD*): This is the minimum TXOP duration that can be assigned to a QSTA and equals the maximum time required to transmit a packet of maximum size for any of the QSTA's TSs. Thus, mTD_i of $QSTA_i$ is calculated as:

$$mTD_i = \max\left(\frac{M_{ij}}{R_{ij}}\right),\ j \in [1,n_i] \qquad (1)$$

where M_{ij} and R_{ij} are the maximum MSDU size and the minimum physical rate for TS_{ij}, respectively.

Maximum TXOP duration (*MTD*): This is the maximum TXOP duration that can be assigned to a QSTA. It should be less than or equal to the transmission time of the **Aggregate Maximum Burst Size (*AMBS*)** of a QSTA. The *AMBS* is the sum of the maximum burst sizes (MBSs) of all TSs of a QSTA. Thus for $QSTA_i$ it holds:

$$AMBS_i = \sum_{j=1}^{n_i} MBS_{ij} \qquad (2)$$

and,

$$MTD_i \le \frac{AMBS_i}{R_i} \qquad (3)$$

where R_i is the minimum physical bit rate assumed for $QSTA_i$ ($R_i=min(R_{ij}),\ j \in [1,n_i]$).

Minimum Service Interval (mSI): It is the minimum time gap required between the start of two successive TXOPs assigned to a specific QSTA. It is calculated as the minimum of the *mSIs* of all the QSTA's TSs:

$$mSI_i = \min\left(mSI_{ij}\right),\ j \in [1,n_i] \qquad (4)$$

If not specified in the TSPEC, mSI_{ij} of TS_{ij} is set equal to the average interval between the generation of two successive MSDUs, i.e., $mSI_{ij}= L_{ij}/\rho_{ij}$.

Maximum Service Interval (MSI): It is the maximum time interval allowed between the start of two successive TXOPs assigned to a QSTA. Although no specific guidelines for calculating *MSI* are provided, an upper limit exists to allow an MSDU generated right after a TXOP assignment to be transmitted at the next TXOP. Accordingly:

$$MSI_i \le D_i - MTD_i \qquad (5)$$

where D_i is defined as the minimum delay bound of all TSs of $QSTA_i$ ($D_i=min(D_{ij}),\ j \in [1,n_i]$). This is an upper limit that ensures that successive TXOPs will be assigned close enough to preserve delay constraints.

2.2 Operation of ARROW Scheduler

Both Simple and SETT-EDD, as briefly described above, decide on TXOP durations using some kind of estimation of the amount of data waiting to be transmitted by every QSTA. ARROW tries to overcome this drawback by adapting TXOP durations based on traffic feedback reports issued by QSTAs The novel characteristic of ARROW is that it exploits the *Queue Size (QS)* field, introduced by 802.11e as part of the new *QoS Data* frames, not supported by legacy 802.11 systems [4], [6], [7]. The QS field can be used by the QSTAs to indicate the amount of buffered traffic for their TSs, i.e., their transmission requirements.

An example of the use of the QS field in ARROW is depicted in Fig. 3. The allocation procedure will be described in detail later in this section. For simplicity reasons, one TS per QSTA is assumed. At time $t_i(x)$, $QSTA_i$ is assigned $TXOP_i(x)$, according to requirements expressed through the QS field of the previous TXOP as well as traffic characteristics and QoS requirements declared in the respective TSPEC. Using a *QoS Data* frame, $QSTA_i$ transmits its data together with the current size of its queue in the QS field ($QS_i(x)$). At time $t_i(x+1)$ the scheduler assigns $TXOP_i(x+1)$ to $QSTA_i$, in order to accommodate the requirements of $QS_i(x)$. During the interval $[t_i(x), t_i(x+1)]$ new data is generated in $QSTA_i$, therefore $QSTA_i$ uses the *QoS Data* frame transmitted at $TXOP_i(x+1)$ to indicate the new queue size ($QS_i(x+1)$). In the same manner, at $t_i(x+2)$ the scheduler assigns $TXOP_i(x+2)$ to $QSTA_i$, accommodating the requirements of $QS_i(x+1)$ and gets the new queue size from $QSTA_i$ ($QS_i(x+2)$). As clearly shown, by utilizing the QS field, ARROW has very accurate information about the time varying properties of each TS, and is able to adapt the TXOP duration accordingly. This is considered essential, especially in the case of bursty and VBR traffic, where transmission requirements feature large time variations.

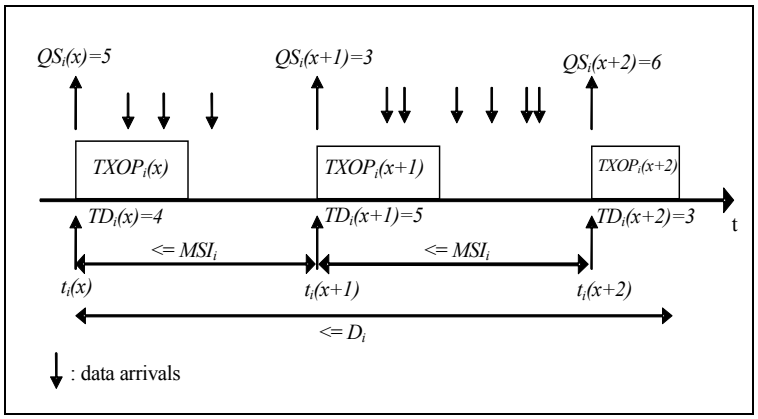


Fig. 3 TXOP assignment with ARROW

As can be observed in Fig. 3, for every $QSTA_i$, data arriving within the interval $[t_i(x), t_i(x+1)]$ can be transmitted no earlier than $TXOP_i(x+2)$ starting at $t_i(x+2)$. Therefore, in order not to exceed the delay deadline of MSDUs, assuming the worst

case that service intervals are equal to MSI_i and $TXOP_i(x+2)=MTD_i$, it should hold that:

$$D_i \geq 2MSI_i + MTD_i \Leftrightarrow$$

$$\Leftrightarrow MSI_i \leq \frac{D_i - MTD_i}{2} \qquad (6)$$

If the scheduler should also take into account possible retransmissions, relation (6) becomes:

$$MSI_i \leq \frac{D_i - MTD_i}{2 + m} \qquad (7)$$

where m is the number of maximum retransmission attempts.

ARROW incorporates a traffic policing mechanism to ensure that the transmission requirements expressed through the QSs do not violate traffic characteristics expressed through the TSPECs. For that purpose, a *TXOP timer* is used, that implements the operation of a leaky bucket of time units. The TXOP timer value T_i for a $QSTA_i$ having n_i active TSs, increases with rate $r(T_i)$:

$$r(T_i) = \sum_{j=1}^{n_i} \left((\frac{L_{ij}}{R_{ij}} + O) / \frac{L_{ij}}{\rho_{ij}} \right) \qquad (8)$$

where O is the overhead due to PHY and MAC headers measured in seconds.

Equation (8) means that during the time interval needed for the generation of an MSDU of Nominal Size at mean data rate, the TXOP Timer should be increased by the time required for the transmission of this MSDU. The maximum TXOP Timer value $\max(T_i)$ equals the time required for the transmission of all maximum bursts:

$$\max(T_i) = \sum_{j=1}^{n_i} \left(\frac{MBS_{ij}}{R_{ij}} + O \right) \qquad (9)$$

According to the operation of ARROW described below, no TXOP longer than the current value of T_i can be assigned to $QSTA_i$ at any time. After each TXOP assignment, the value of the respective TXOP timer is reduced accordingly.

The operation of ARROW can be divided in the following steps:

1. The scheduler waits for the channel to become idle.
2. When the channel becomes idle at a given moment t, the scheduler checks for QSTAs that:

   a. can be polled without violating *mSI*, i.e., for a $QSTA_i$ that was last polled at time t_i, it should hold that:

$$t \geq t_i + mSI_i \qquad (10)$$

   and,

   b. their TXOP timer value T is greater than the value of their mTD, to ensure enough time for the minimum TXOP duration.
3. If no QSTAs are found, the scheduler waits until (10) becomes true at least for one QSTA and returns to step 2.
4. In different case, the scheduler polls the QSTA with the earliest deadline. The deadline for a $QSTA_i$ is the latest time that this QSTA should be polled, i.e., t_i+MSI_i, where t_i is the time of the last poll for $QSTA_i$.

5. Assuming $QSTA_i$ having n_i active TSs is selected for polling, the scheduler calculates TD_i, as follows:

a. For every TS_{ij} of $QSTA_i$ ($j \in [1,n_i]$), the scheduler calculates TD_{ij}, as the maximum of (i) the time required to accommodate the pending traffic, as indicated by the queue size of that TS (QS_{ij}), plus any overheads (O), and, (ii) mTD_{ij}, to ensure that the assigned TXOP will have at least the minimum duration:

$$TD_{ij} = \max\left(\frac{QS_{ij}}{R_{ij}} + O, mTD_{ij}\right) \qquad (11)$$

In the special case where QS_{ij} is equal to zero, TD_{ij} is set equal to the time for the transmission of a Null-Data MSDU. In this way, $QSTA_i$ is allowed to transmit a Null-Data MSDU, in order to update the queue size information for TS_{ij}. TD_i for $QSTA_i$ is calculated as the sum of all TD_{ij}:

$$TD_i = \sum_{j=1}^{n_i} TD_{ij} \qquad (12)$$

b. Finally TD_i obtained from (12) is compared with the current TXOP Timer value T_i, to ensure conformance with the negotiated traffic profile:

$$TD_i = \min(TD_i, T_i) \qquad (13)$$

6. After the scheduler assigns the TXOP, it reduces the respective TXOP timer value accordingly and returns to step 1:

$$T_i = T_i - TD_i \qquad (14)$$

2.2 Simulation results

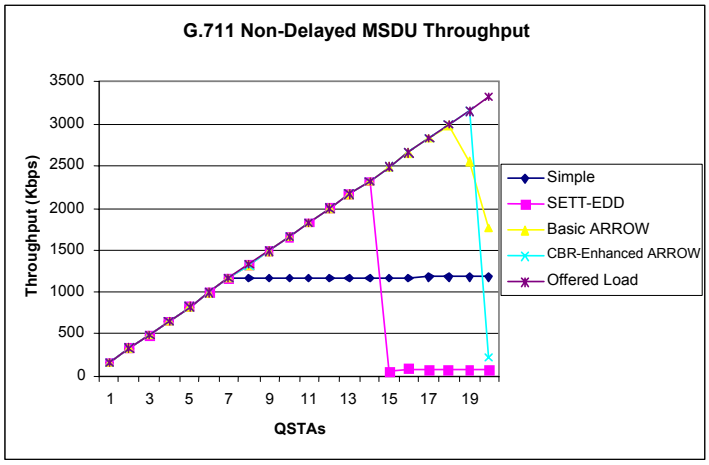
To measure the performance ARROW against Simple and SETT-EDD, a specialized 802.11e simulation tool developed by ATMEL Hellas was used [8]. The simulation scenarios considered an increasing number of QSTAs attached to a QAP. All QSTAs and the QAP were supporting the extended MAC layer specified in IEEE 802.11e and the PHY layer specified in IEEE 802.11g, with a transmission rate of 12Mbps. Each QSTA had two active sessions:

a. a bi-directional G.711 voice session (CBR traffic), mapped into two TSs (one per direction), and,

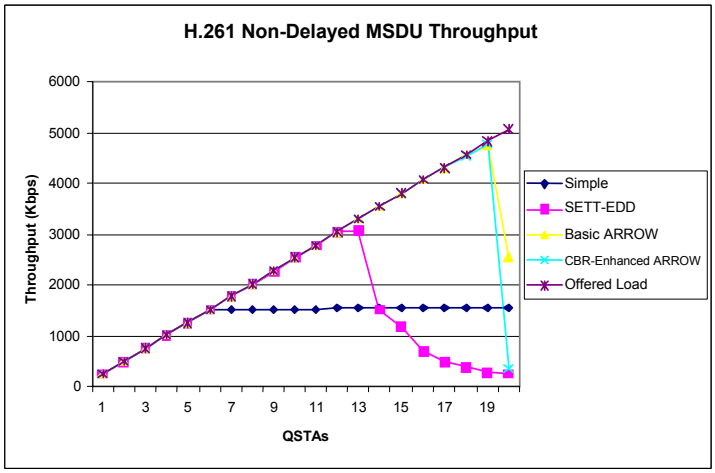
b. an uplink (from QSTA to QAP) H.261 video session at 256 Kbps (VBR traffic), mapped into one uplink TS.

Fig. 4 depicts throughput of non-delayed MSDUs for voice and video traffic. For voice traffic (Fig. 4a), basic ARROW accommodates up to 18 QSTAs, while SETT-EDD can manage up to 14 QSTAs and Simple up to only 7 QSTAs. Using the enhancement for CBR traffic, the number of QSTAs can be increased to 19 with CBR-enhanced ARROW, as a result of less required overhead. For video traffic (Fig. 4b), basic and CBR-enhanced ARROW outperform both SETT-EDD and Simple, accommodating up to 19 QSTAs, as opposed to 13 with SETT-EDD and 6 with Simple. The main reason for the considerably improved performance of basic ARROW is the accurate TXOP assignment it performs, by utilizing the queue size

information. This is also shown in detail using more metrics later in this section. From Fig 4a and 4b it is clear that CBR-enhanced ARROW can extend the admission capability of the system, as it can accommodate up to 19 QSTAs with voice and video TSs.



(a) G.711 Voice



(b) H.261 video

Figure 2. Throughput of Non-Delayed MSDUs

It is interesting to observe that throughput of SETT-EDD and ARROW (both basic and CBR-enhanced) reduces rapidly immediately after reaching its maximum value. The reason is that, due to the dynamic TXOP assignment performed by these algorithms, new TSs entering the system can participate equally to the channel assignment. This means that, after the overall input traffic exceeds a value that

corresponds to the maximum capability of the scheduler, none of the TSs (new or old) is serviced as required. The Simple Scheduler on the other hand, manages to provide a stable throughput regardless of the offered load, because static allocations for existing TSs are not affected by the traffic load increase. This effect highlights the need for an effective admission control scheme for SETT-EDD and ARROW that would prevent the offered load from exceeding the maximum scheduling capability.

References

1. D.Skyrianoglou, N.Passas, A. K. Salkintzis, E. Zervas “A Generic Adaptation Layer for Differentiated Services and Improved Performance in Wireless Networks”, in Proc. PIMRC 2002, Lisbon, Portugal, September 2002.
2. D.Skyrianoglou, N.Passas, L.Merakos, “Improving QoS in a Multi-Hop Wireless Environment”, in Proc. Med-Hoc-Net 2002, Sardegna, Italy, September 2002.
3. D.Skyrianoglou, N.Passas and S.Kampouridou, “A DiffServ-based Classification Scheme for Internet Traffic Over Wireless Links”, in Proc. ICWLHN 2001, Singapore, December 2001.
4. D.Skyrianoglou, N.Passas and A.Salkintzis, “ARROW: An Efficient Traffic Scheduling Algorithm for IEEE 802.11e HCCA”, IEEE Transactions on Wireless Communications, vol. 5, no.12, December 2006.
5. N. Passas, D. Skyrianoglou, and P. Mouziouras, “Prioritized Support of Different Traffic Classes in IEEE 802.11e Wireless LANs”, Elsevier Computer Communications Journal, vol. 29, no. 15, September 2006.
6. D.Skyrianoglou, N.Passas and A.Salkintzis, “Traffic Scheduling for Multimedia QoS Over Wireless LANs”, in Proc. ICC 2005, Seoul, Korea, May 2005.
7. D.Skyrianoglou, N.Passas and A.Salkintzis, "Traffic Scheduling in IEEE 802.11e Networks Based on Actual Requirements", in Proc. Mobile Venue 2004, Athens, Greece, May 2004.
8. A.Floros, D. Skyrianoglou, N. Passas, T. Karoubalis, “A Simulation Platform for QoS Performance Evaluation of IEEE 802.11e ”, The Mediterranean Journal of Computers and Networks, vol. 2, no. 2, April 2006.
9. A. Salkintzis, N. Passas, and D. Skyrianoglou, “On the Support of Voice Call Continuity across UMTS and Wireless LANs”, Wiley Wireless Communications and Mobile Computing (WCMC) Journal, vol. 8, issue 7, Sep. 2008.
10. A. Salkintzis, N. Passas, and D. Skyrianoglou, “Seamless Voice Call Continuity in 3G and WLANs”, in Proc. 9th International Symposium on Wireless Personal Multimedia Communications (WPMC), San Diego, CA, September 2006.
11. A. Salkintzis, G. Dimitriadis, D. Skyrianoglou, N. Passas, F.-N. Pavlidou, “Seamless Continuity of Real-Time Video Across UMTS and WLAN Networks: Challenges and Performance Evaluation”, Special issue on “Towards Seamless Interworking of WLAN and Cellular Networks”, IEEE Wireless Communications Magazine, vol. 12, no. 3, Jun. 2005.
12. A Salkintzis, D. Skyrianoglou and N. Passas, "Seamless Multimedia QoS Across UMTS and WLAN Networks", in Proc. IEEE Vehicular Technology Conference (VTC) Spring 2005, Stockholm, Sweden, May 2005.
13. D. Skyrianoglou, N. Passas and A. Salkintzis, “Support of IP QoS over Wireless LANs”, in Proc. VTC Spring ’04, Milan, Italy, May 2004.
14. D. Skyrianoglou and N. Passas, “A Framework for Unified IP QoS Support Over UMTS and Wireless LANs”, in Proc. European Wireless 2004, Barcelona, Spain, February 2004.
15. N. Passas, D. Skyrianoglou and A. Salkintzis, "Supporting UMTS QoS in WLANs" in Proc. Personal Wireless Communications (PWC) 2003, Venice, Italy, September 2003.

Optical Process and Analysis of Historical Documents

Nikolaos Stamatopoulos*

[1] Department of Informatics and Telecommunications
National and Kapodistrian University of Athens

[2] Computational Intelligence Laboratory
Institute of Informatics and Telecommunications
National Centre for Scientific Research "Demokritos"
nstam@iit.demokritos.gr

Abstract. The collections of historical books are an important source of information, both for the history of previous periods and for the development of the cultural documentation itself. Although to date, there have been made several attempts of digitalization and electronic navigation, there is not an appropriate frame of optical process and analysis of the content of these collections, consequently a large number of historical books have not been studied yet and remain unexploited. In this thesis, we studied the preprocessing stages which are performed before the recognition process and we focused on the enhancement and segmentation of historical documents. Preprocessing stages play an important role in document image processing since they affect the performance of subsequent processing, such as optical character recognition. At the enhancement stage, we focused on the border removal as well as on the dewarping of document images, which are common problems associated with historical documents. Two methodologies that detect and remove black borders as well as noisy text regions are proposed. Furthermore, optimal page frames of double page document images are detected. The experimental results on several historical documents demonstrate the effectiveness of the proposed techniques. Concerning the warping problem, a coarse-to-fine rectification methodology to compensate for undesirable document image distortions is proposed. To verify the validity of the proposed methodology, experiments have been carried out using indirect evaluation techniques as well as a novel semi-automatic evaluation methodology. At the document image segmentation stage we proposed a novel combination method of complementary text line segmentation techniques. Furthermore, a methodology for character segmentation in historical documents is suggested. Comparative experiments using several historical documents from different languages and time periods prove the efficiency of the proposed technique. Finally, in order to ease the construction of document image segmentation ground-truth that includes text-image alignment we presented an efficient technique.

Keywords: document image enhancement, border removal, document image dewarping, document image segmentation, combined segmentation techniques

* Dissertation Advisors: [1] Sergios Theodoridis, Professor – [2] Basilis Gatos, Researcher

1 Introduction

Recognition of historical documents is essential for quick and efficient content exploitation of the valuable historical collections that are part of our culture heritage. Several factors such as low paper quality, dense and arbitrary layout, low print contrast, typesetting imperfections, lack of standard alphabets and fonts do not permit the application of conversional recognition techniques to historical documents. Due to these reasons, recognition of historical documents is one of the most challenging tasks in document image processing.

In this thesis, we studied the preprocessing stages which are performed before the recognition process and we focused on the enhancement and segmentation of historical documents. Preprocessing stages play an important role in document image processing since they affect the performance of subsequent processing, such as optical character recognition. At the enhancement stage, we focused on the border removal as well as on the dewarping of document images, which are common problems associated with historical documents. Moreover, at the document image segmentation stage we proposed a novel combination method of complementary text line segmentation technique as well as a methodology for character segmentation in historical documents. Finally, in order to ease the construction of document image segmentation ground-truth that includes text-image alignment we presented an efficient technique.

2 Document Image Enhancement

2.1 Border Removal

Document images are often framed by a noisy black border or include noisy text regions from neighbouring pages when captured by a digital camera. Approaches proposed for document segmentation and character recognition usually consider ideal images without noise. However, there are many factors that may generate imperfect document images. When a page of a book is captured by a camera, text from an adjacent page may also be captured into the current page image. These unwanted regions are called “noisy text regions”. Additionally, there will usually be black borders in the image. These unwanted regions are called “noisy black borders”. All these problems influence the performance of segmentation and recognition processes. There are only few techniques in the literature for page borders detection [1-5]. Most of them detect only noisy black borders and not noisy text regions.

We propose a new and efficient algorithm for detecting and removing noisy black borders as well as noisy text regions [6]. This algorithm uses projection profiles and a connected component labelling process to detect page borders. Additionally, signal cross-correlation is used in order to verify the detected noisy text areas. The experimental results on several historical document images indicate the effectiveness of the proposed technique.

Moreover, document images are usually produced by scanning books or periodicals. Scanning two pages at the same time is a very common practice as it

helps to accelerate the scanning process. However, it may affect the performance of subsequent processing such as document analysis and optical character recognition (OCR) since the majority of approaches are able to process only single page images. Furthermore, another drawback of scanning two pages at the same time is the appearance of noisy black borders around text areas as well as of noisy black stripes between the two pages. For these reason, we propose a novel methodology that detects the optimal page frames of double page document images that is based on the vertical and horizontal white run projections [7]. Our aim is to split the image into the two pages as well as to remove noisy borders. At a first step, a pre-processing which includes binarization, noise removal and image smoothing is applied. At a next step, the vertical zones of the two pages are detected. Finally, the frame of both pages is detected after calculating the horizontal zones for each page.

2.2 Dewarping

Document image acquisition by a flatbed scanner or a digital camera often results in several unavoidable image distortions due to the form of printed material (e.g. bounded volumes), the camera setup or environmental conditions (e.g. humidity that causes page shrinking). Text distortions not only reduce document readability but also affect the performance of subsequent processing such as document layout analysis and optical character recognition (OCR).

Over the last decade, many different techniques have been proposed for document image rectification and they can be classified into two main categories based on (i) 3D document shape reconstruction [8-9] and (ii) 2D document image processing [10-15]. Techniques of the former category obtain the 3D information of the document image using special setup or reconstruct the 3D model from information existing in document images. On the other hand, techniques in the latter category do not depend on auxiliary hardware or prior information but they only rely on 2D

In this thesis, we propose a goal-oriented rectification methodology to compensate for undesirable distortions of document images captured by flatbed scanners or hand-held digital cameras (TSD.ver2) [16]. The proposed technique is directly applied to the 2D image space without any dependence to auxiliary hardware or prior information. It first detects words and text lines to rectify the document image in a coarse scale and then further normalize individual words in finer detail using baseline correction. Although the coarse rectification stage applies word and text line detection at the original distorted document image, which is a well-known hard task, potential erroneous detection results do not seriously affect it as it requires only some specific points. Experimental results on several document images with a variety of distortions show that the proposed method produces rectified images that give a significant boost in OCR performance. This work is an extension of our previous work (TSD.ver1) [17] which incorporates a new method for the curved surface projection, the word baseline fitting as well as the restoration of horizontal alignment. We also propose to rectify the distortion of individual words using baseline estimation. Finally, we propose a new semi-automatic evaluation method [18] based on matching manually marked points of the original image and corresponding points of the rectified image. A quantitative measure is calculated to evaluate the performance of our method.

3 Document Image Segmentation

3.1 Text Line Segmentation

In document analysis and recognition, several approaches have been proposed for improving OCR accuracy through combination [19]. These approaches can be categorized in two categories: (i) techniques in classifier combinations and (ii) string alignment combination methods [20]. Approaches of the second category combine several OCR outputs to produce a more accurate string estimate of the original text, but this cannot be done on character-by-character basis because of segmentations errors. Outputs strings must be aligned to extract an estimate and also errors must be uncorrelated.

Based on a similar way of thought we could combine the results of different segmentation techniques in order to achieve better segmentation results. Document segmentation into text lines is a major task in a document image analysis system. A wide variety of methods have been proposed in the literature for document segmentation which can be categorized in five major categories: (1) projection profiles methods; (2) smearing methods; (3) methods based on the Hough transform; (4) grouping methods and (5) stochastic methods. Techniques from each category can confront some specific problems such as overlapping, touching components, image degradations, variability in skew angles and directions, disturbing elements, variability in inter-word and inter-character distances and others. So, we propose a combination method of complementary segmentation techniques where each technique can solve some different difficult problems [21]. Our goal is to increase the efficiency and the accuracy of the segmentation result using (i) the results of segmentation techniques which belong to different categories and (ii) specific features of the initial document.

3.2 Character Segmentation

The most recognition errors are due to character segmentation errors. Very often, even in printed text, adjacent characters are touching, and may exist in an overlapped field. Therefore, it is essential to segment a given word correctly into its character components. Any failure or error in this segmentation step can lead to a critical loss of information from the document. Character segmentation previous work concerns mostly handwritten text but methods for machine-printed text have also been proposed [22-23].

The proposed character segmentation algorithm [24] is based on skeleton segmentation paths which are used to isolate possible connected characters. The basic idea is that we can find possible segmentation paths linking the feature points on the skeleton of the word and its background

3.3 Creation of Document Image Segmentation Ground Truth

Efficient ground truth creation is essential for training and evaluation purposes in the document image analysis and recognition pipeline. Since a large number of tools have to be trained and evaluated in realistic circumstances we need to have a quick and low cost way to create the corresponding ground truth. Moreover, the specific need for having the correct text correlated with the corresponding image area in text line and word level makes the process of ground truth creation a difficult, tedious and costly task. Transcript mapping (or text alignment) techniques are used in order to map the correct text information to a segmentation result produced automatically. Usually, these techniques are very useful in order to automatically create benchmarking data sets. They are mainly based on hidden Markov models (HMMs) [25] and dynamic time warping (DTW) [26] and mainly focus on the alignment of handwritten document images with the corresponding transcription on word level.

We introduce an efficient transcript mapping technique to ease the construction of document image segmentation ground truth that includes text-image alignment in text line, word and character level [27]. We facilitate the annotation of text line, word and character segmentation ground truth regions as well as the correlation with corresponding text making use of the correct document transcription. In the proposed framework, we assume that the transcription includes the correct text line break information. This information is used in a novel transcript mapping module in order to efficiently create the text line, word and word segmentation ground truth. The proposed text line transcript mapping technique is based on Hough transform that is guided by the number of the text lines in order to efficiently create the text line segmentation result. Concerning the word and character segmentation ground truth, a gap classification technique constrained by the number of the words and character is used. We recorded that using the proposed technique for handwritten documents, the percentage of time saved for ground truth creation and text-image alignment is more than 90%.

4 Experimental Results

4.1 Border Removal

The performance evaluation method used is based on a pixel based approach and counts the pixels at the correct page frames and the detected page frames. For this purpose, we manually mark the correct page frames in the original document image in order to create the ground truth set. Let G be the set of all pixels inside the correct page frame in ground truth, R the set of all pixels inside the result page frame and T(s) a function that counts the elements of set s. We calculate the Precision and Recall as follows:

$$Precision = \frac{T(G \cap R)}{T(R)} \;\&\; Recall = \frac{T(G \cap R)}{T(G)} \tag{1}$$

A performance metric *FM* can be extracted if we combine the values of precision and recall:

$$FM = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (2)$$

To verify the validity of the proposed method [6] we used two different datasets. The first ("POLYTIMO") was a set of Greek historical documents [28] consisted of 370 document images. The second set ("IMPACT") [29] consisted of 22383 historical documents including newspapers, periodical etc. For comparison purposes, we applied at the same dataset the state-of-the-art method [1] as well as the commercial packages BookRestorer [30], WiseBook [31] and ScanFix [32]. Tables 1 and 2 illustrate the overall evaluation results.

Table 1. Border Removal - Evaluation Results using "POLYTIMO" dataset.

Method	Precision (%)	Recall (%)	F-Measure (%)
Proposed Method [6]	91.11	96.95	93.94
Le et al. [1]	70.90	99.33	82.74
BookRestorer [30]	74.33	95.47	83.58
WiseBook [31]	53.00	99.02	69.05
ScanFix [32]	51.49	87.96	64.95

Table 2. Border Removal - Evaluation Results using "IMAPCT" dataset.

Method	Precision (%)	Recall (%)	F-Measure (%)
Proposed Method [6]	98.62	98.46	98.54
Le et al. [1]	97.28	94.01	95.62
BookRestorer [30]	94.11	96.92	95.50
WiseBook [31]	83.32	98.94	90.46
ScanFix [32]	85.00	98.04	91.05

To verify the validity of the proposed method [7] we used 3467 double page document images from 50 different historical books. For comparison purposes, we applied at the same dataset the commercial package ABBYY FineReader Engine 10 [33]. Table 3 illustrates the overall evaluation results.

Table 3. Border Removal & Page Split - Evaluation Results

Method	Precision (%)	Recall (%)	F-Measure (%)
Proposed Method [7]	92.04	98.35	95.09
ABBYY FineReader Engine 10 [33]	62.66	91.53	74.39

4.2 Dewarping

To verify the validity of the proposed methodology we use as a performance measure the character and word accuracy metrics by carrying out OCR on original and rectified document images. Furthermore, experiments have been carried out using the

proposed semi-automatic evaluation methodology [18]. The experimental results from both procedures are presented in the sequel.

OCR Evaluation: The use of OCR as a means for indirect evaluation is widely used in the evaluation of rectification techniques. Character accuracy metric is defined as the ratio of the number of correct characters (number of characters in the correct document transcription minus the number of errors) over the total number of characters in the correct document transcription:

$$Character\ Accuracy = (\#chars - \#errors) / \#chars \quad (3)$$

In order to define the errors we count the minimum number of edit operations (insertion, deletion or substitution) that are required to correct the text generated by the OCR system (string edit distance). Moreover, we carried out OCR testing on original and rectified document images using also the word accuracy metric. Word accuracy is defined as the ratio of the number of correct words (number of words in the correct document transcription minus number of misrecognized words) to the total number of word in the correct document transcription:

$$Word\ Accuracy = (\#words - \#misrecognized\_words) / \#words \quad (4)$$

We used a dataset of 100 distorted document images at 200 dpi. The document images contain different font sizes and suffer from several distortions. For comparison purposes, we applied at the same dataset the first version of the proposed method [17], the state-of-the-art method [14] as well as the commercial package BookRestorer [30]. OCR testing is performed using ABBYY FineReader Engine 8.1 [33]. Both the distorted document images and the rectified documents are fed into OCR Engine for text recognition. Table 4 illustrates the average character accuracy as well as the average word accuracy.

Table 4. Average Character and Word Accuracy on 100 Document Images

Rectification Technique	#characters	Character Accuracy	#words	Word Accuracy
Without Rectification	170726	56,54%	27012	44,78%
Gatos et. al [14]	170726	81,51%	27012	62,71%
Proposed method TSD.ver1	170726	85,56%	27012	66,06%
BookRestorer [30]	170726	90,52%	27012	78,85%
Proposed method TSD.ver2	170726	93,82%	27012	84,07%

Semi-Automatic Evaluation: The evaluation methodology proposed in [18] avoids the dependence on an OCR engine or human interference. It is based on a point-to-point matching procedure using Scale Invariant Feature Transform (SIFT) [34] as well as the use of cubic polynomial curves for the calculation of a comprehensive measure which reflects the entire performance of a rectification technique in a concise quantitative manner. First, the user manually mark specific points on the distorted document image which correspond to N appropriate text lines of the document with

representative deformation. Then, using SIFT transform, the marked points of the distorted document image are matched to the corresponding points of rectified document image. Finally, the cubic polynomial curves which fit to these points are estimated and are taken into account in the evaluation measure DW:

$$\boldsymbol{DW} = \frac{\sum_{j=1}^{N} DW_j}{N} \times 100\% \tag{5}$$

where DW_j is the measure which reflects the performance of the rectification technique with respect to the j^{th} selected text line. DW_j equals to one when the j^{th} selected text line in the rectified document image is a horizontal straight text line that is the expected optimal result. It shows that the rectification technique produces the best result. On the other hand, DW_j equals to zero when the rectified document image is equal to or worse than the original image. Therefore, DW ranges in the interval [0,…,100] and the higher the value of DW, the better is the performance of the rectification technique. Table 5 illustrates the average DW measure of all rectification methods. It is worth mentioning that the overall comparative ranking is the same with the one which is produced with the experiment that takes into account OCR performance. The proposed rectification method outperforms all the others methods.

Table 5. Comparative Results Using the Semi-Automatic Evaluation Methodology

Rectification Technique	**DM**
Gatos et. al [14]	79.35
Proposed method TSD.ver1	82.53
BookRestorer [30]	84.12
Proposed method TSD.ver2	91.90

4.3 Text Line Segmentation

To verify the validity of the proposed method we use two complementary line segmentation methods, projection profiles based on [35] and Adaptive RLSA based on [24]. In [35], each minimum of the profile curve is a potential segmentation point. Potential points are then scored according to their distance to adjacent segmentation points. The reference distance is obtained from the histogram of distances between adjacent potential segmentation points. The highest scored segmentation point is used as an anchor to derive the remaining ones. In [24], Makridis et. al propose the adaptive RLSA which is an extension of the classical RLSA in the sense that additional smoothing constraints are set in regard to the geometrical properties of neighbouring connected components. The replacement of background pixels with foreground pixels is performed when these constraints are satisfied.

We apply each method to a set of 50 historical documents images (1633 text line segments) as well as to a set of 50 handwritten documents (1144 text line segments). Then, using the two different segmentation results for each image, we generate a new segmentation result according to the proposed combination method [21].

For the purpose of the evaluation, we manually marked the correct line segments in the document images. The performance evaluation was based on counting the number of matches between the lines detected by the segmentation algorithms or their combination and the lines in the ground truth [36]. Finally, we calculate the detection rate (DR), the recognition accuracy (RA) as well as the F-Measure (FM). As depicted in Tables 6 and 7, the new segmentation result outperforms the two others methods and it increases the overall evaluation measure about 20%.

Table 6. Comparative Results - Historical Document Images

Segmentation Technique	GT regions	Result regions	One-to-one matches	DR (%)	RA (%)	FM (%)
Projection Profile [35]	1633	1577	1327	81.26	84.15	82.68
Adaptive RLSA [24]	1633	1594	1358	83.16	85.19	84.16
After combination using the proposed method [21]	1633	1605	1529	93.63	95.26	94.44

Table 7. Comparative Results - Handwritten Document Images

Segmentation Technique	GT regions	Result regions	One-to-one matches	DR (%)	RA (%)	FM (%)
Projection Profile [35]	1144	1248	841	73.51	67.39	70.32
Adaptive RLSA [24]	1144	1314	860	75.17	65.45	69.98
After combination using the proposed method [21]	1144	1152	1071	93.62	92.97	93.29

4.4 Character Segmentation

In order to record the efficiency of the proposed character segmentation method we followed a well established evaluation approach that is also employed by several document segmentation contests. The performance evaluation method is based on counting the number of matches between the entities detected by the algorithm and the entities in the ground truth. Finally, we calculate the detection rate (DR), the recognition accuracy (RA) as well as the F-Measure (FM). We used a set of 51 historical document images and compared with the commercial products ABBYY FineReader Engine 8.1 [33] and with the open source OCRopus libray [37] as well as with two state-of-the-art methods based on RLSA [22] and on Projection Profiles [23]. Table 8 presents the evaluation results.

Table 8. Character Segmentation - Evaluation Results

Segmentation method	GT regions	Result regions	One-to-one matches	DR (%)	RA (%)	FM (%)
Projection Profiles [23]	71818	71948	49449	68.85	68.73	68.79
RLSA [22]	71818	69065	56361	78.48	81.61	80.01
ABBYY FineReader Engine 8.1 [33]	71818	74721	52782	73.49	70.64	72.04
OCROpus [37]	71818	79575	53648	74.70	67.42	70.87
Proposed method [24]	71818	75955	62425	86.92	82.19	84.49

5 Concluding Remarks

In this thesis, we studied the preprocessing stages which are performed before the recognition process and we focused on the enhancement and segmentation of historical documents. Preprocessing stages play an important role in document image processing since they affect the performance of subsequent processing, such as optical character recognition. At the enhancement stage, we focused on the border removal as well as on the dewarping of document images, which are common problems associated with historical documents. Two methodologies that detect and remove black borders as well as noisy text regions are proposed. Furthermore, optimal page frames of double page document images are detected. Concerning the warping problem, a coarse-to-fine rectification methodology to compensate for undesirable document image distortions is proposed. To verify the validity of the proposed methodology, experiments have been carried out using indirect evaluation techniques as well as a novel semi-automatic evaluation methodology. At the document image segmentation stage we proposed a novel combination method of complementary text line segmentation techniques. Furthermore, a methodology for character segmentation in historical documents is suggested. Finally, in order to ease the construction of document image segmentation ground-truth that includes text-image alignment we presented an efficient technique.

References

1. D.X. Le, G.R. Thoma and H. Wechsler, “Automated borders detection and adaptive segmentation for binary document images”, International Conference on Pattern Recognition, Vienna, Austria, 1996, pp. 737-741.
2. C. Fan, Y.K. Wang and T.R. Lay, “Marginal noise removal of document images”, Pattern Recognition, vol. 35, no. 11, 2002, pp. 2593-2611.

3. B.T. Avila and R.D. Lins, “A New Algorithm for Removing Noisy Borders from Monochromatic Documents”, ACM Symposium on Applied Computing, Nicosia, Cyprus, 2004, pp. 1219-1225.
4. B.T. Avila and R.D. Lins, “Efficient Removal of Noisy Borders from Monochromatic Documents”, International Conference on Image Analysis and Recognition, Porto, Portugal, 2004, pp. 249-256.
5. F. Shafait, J.v. Beusekom, D. Keysers and T.M. Breuel, “Document cleanup using page frame detection”, International Journal on Document Analysis and Recognition, vol. 11, no. 2, 2008, pp. 81-96.
6. N. Stamatopoulos, B. Gatos and A. Kesidis, “Automatic Borders Detection of Camera Document Images”, 2nd International Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 71-78.
7. N. Stamatopoulos, B. Gatos and T. Georgiou, “Page Frame Detection for Double Page Document Images”, 9th International Workshop on Document Analysis Systems, Boston, MA, USA, 2010, pp. 401-408.
8. M.S. Brown and W.B. Seales, “Image restoration of arbitrarily warped documents”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 10, 2004, pp. 1295-1306.
9. C.L. Tan, L. Zhang, Z. Zhang and T. Xia, “Restoring warped document images through 3D shape modeling”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 2, 2006, pp. 195-208.
10. L. Zhang and C.L. Tan, “Warped image restoration with applications to digital libraries”, 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 192-196.
11. H. Ezaki, S. Uchida, A. Asano and H. Sakoe, “Dewarping of document image by global optimization”, 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 302-306.
12. A. Ulges, C.H. Lampert and T.M. Breuel, “Document image dewarping using robust estimation of curled text lines”, 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 1001-1005.
13. S.J. Lu, B.M. Chen and C.C. Ko, “A partition approach for the restoration of camera images of planar and curled document”, Image and Vision Computing, vol. 24, no. 8, 2006, pp. 837–848.
14. B. Gatos, I. Pratikakis and K. Ntirogiannis, “Segmentation Based Recovery of Arbitrarily Warped Document Images”, 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 989-993.
15. M.S. Brown and Y.C. Tsoi, “Geometric and shading correction for images of printed materials using boundary”, IEEE Transactions on Image Processing, vol. 15, no. 6, 2006, pp. 1544-1554.
16. N. Stamatopoulos, B. Gatos, I. Pratikakis and S.J. Perantonis, “Goal-oriented Rectification of Camera-Based Document Images”, IEEE Transactions on Image Processing, vol. 20, no. 4, 2011, pp. 910-920.
17. N. Stamatopoulos, B. Gatos, I. Pratikakis and S.J. Perantonis, “A Two-Step Dewarping of Camera Document Images”, 8th International Workshop on Document Analysis Systems Nara, Japan, 2008, pp. 209-216.
18. N. Stamatopoulos, B. Gatos and I. Pratikakis, “A Methodology for Document Image Dewarping Techniques Performance Evaluation”, 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 956-960.
19. J.C. Handley, "Improving OCR accuracy through combination: a survey", International Conference on Systems, Man, and Cybernetics, California, USA, 1998. pp. 4330-4333.

20. S. V. Rice, J. Kanai, T. A. Nartker, "An Algorithm for Matching OCR-Generated Text Strings", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 8, Issue 5, 1994, pp. 1259-1268.
21. N. Stamatopoulos, B. Gatos and S.J. Perantonis, "A Method for Combining Complementary Techniques for Document Image Segmentation", Pattern Recognition, vol. 42, no. 12, 2009, pp. 3158-3168.
22. K. Khurshidt, "Analysis and Retrieval of Historical Document Images", PhD Thesis, Université Paris Descartes, 2009.
23. A. Antonacopoulos and D. Karatzas, "Semantics-based content extraction in typewritten historical documents", 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 48–53.
24. N. Nikolaou, M. Makridis, B. Gatos, N. Stamatopoulos and N. Papamarkos, "Segmentation of historical machine-printed documents using Adaptive Run Length Smoothing and skeleton segmentation paths", Image and Vision Computing, vol. 28, no. 4, 2010, pp. 590-604.
25. A. Toselli, V. Romero, E. Vidal, "Viterbi based alignment between text images and their transcripts" Workshop on Language Technology for Cultural Heritage Data, 2007, pp.9-16.
26. C.V. Jawahar, A. Kumar, "Content-level Annotation of Large Collection of Printed Document Images" Int. Conference on Document Analysis and Recognition, 2007, pp.799-803.
27. N. Stamatopoulos, G. Louloudis and B. Gatos, "A Comprehensive Evaluation Methodology for Noisy Historical Document Recognition Techniques", 3rd Workshop on Analytics for Noisy Unstructured Text Data, Barcelona, Spain, 2009, pp. 47-54.
28. POLYTIMO project, http://iit.demokritos.gr/cil/POLYTIMO.
29. IMPACT project, European Community's Seventh Framework Programme under grant agreement N° 215064, http://www.impact-project.eu/
30. BookRestorer: http://www.i2s-bookscanner.com/
31. WiseBook: http://www.cadcam.org/wise-book.php
32. ScanFix Xpress: http://www.accusoft.com/scanfix.html
33. ABBYY FineReader: http://finereader.abbyy.com/
34. G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, vol. 60, no. 2, 2004, pp. 91-110.
35. A. Antonacopoulos, D. Karatzas, "Document Image analysis for World War II personal records", First International Workshop on Document Image Analysis for Libraries, Palo Alto, 2004, pp. 336-341.
36. B. Gatos, N. Stamatopoulos and G. Louloudis, "ICDAR2009 Handwriting Segmentation Contest", International Journal on Document Analysis and Recognition, vol. 14, no. 1, 2011, pp. 25-33.
37. The OCROpus open source document analysis system: http://code.google.com/p/ocropus/

Efficient algorithms for the study and design of parallel robotic mechanisms with physiotherapy applications

Christos E. Syrseloudis*
National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
15784 Athens Greece
chsirsel@di.uoa.gr

Abstract

The aim of this work is to study the design and identification of kinematics parameters of robotic mechanisms for ankle physiotherapy applications. We begin with the study of the ankle joint complex and we adopt the 2 axes kinematics model from the literature which provides the necessary DOFs and the kind of the ankle motions. The objective is to describe the basic movements of the foot about the ankle joint. We accomplish the design requirements of a physiotherapy robot by performing appropriate experimental measurements. These requirements are the necessary workspace, velocities, accelerations and torque bounds. We examine two existing robotic architectures while we finally introduce a new 2-DOF hybrid serial-parallel robot with mechanical adjustability as an ankle physiotherapy device. The advantages of this physiotherapy device against the existing physiotherapy robots are the minimum number of its actuators, its increased safety, modularity and economy. Then we perform the parametric design of this platform which has been based on the predefined design specifications and then we evaluate the design via simulations. Finally, we develop a simple, accurate and robust identification method of the kinematics parameters of the ankle joint complex. This method combines the concept of robot calibration and arc trajectory fitting in 3D circles.

1 Introduction

In this dissertation we focuss on the design and identification of kinematics parameters of robotic mechanisms for ankle physiotherapy applications. Our contribution is in the following three main topics:
-*Design framework*: A unified design framework for the design of an ankle rehabilitation robot is missing from the literature. We study the ankle kinematics modeling and we complete with experimental data a set of design requirements for an ankle rehabilitation robot.
-*Design of a new ankle rehabilitation robot*: The existing ankle rehabilitation robots are either redundant or not follow exactly the ankle movements. The detailed study of the ankle kinematics led us to introduce a new hybrid parallel-serial robot with 2-DOF and mechanical adjustability. The parametric design of the new robot is carried out relying on the design specification described above. This physiotherapy device outperforms to the existing physiotherapy robots on the minimum number of actuators, safer movements, modularity and economy.
-*Identification of the ankle joint complex kinematics*: The different kinematics characteristics between patients reveal the need for identification of the ankle parameters for the appropriate tuning of a physiotherapy robot. We develop a simple and robust method for identification of the ankle joint kinematics which combines the concept of robot calibration and arc trajectory fitting in 3D circles. Despite to the existing identification methods, our method avoids the use of position tracking of multiple point-markers on the body-member and the use of expensive optical motion analysis systems. This makes it applicable in a physiotherapy clinic.

1.1 Previous Work

Our study begins with the structure and kinematics modeling of the ankle joint complex. The human ankle has a complex multi-joint structure which determines the motion of the foot with respect to the shank. A survey in [7] about the ankle-joint complex modeling starts with early

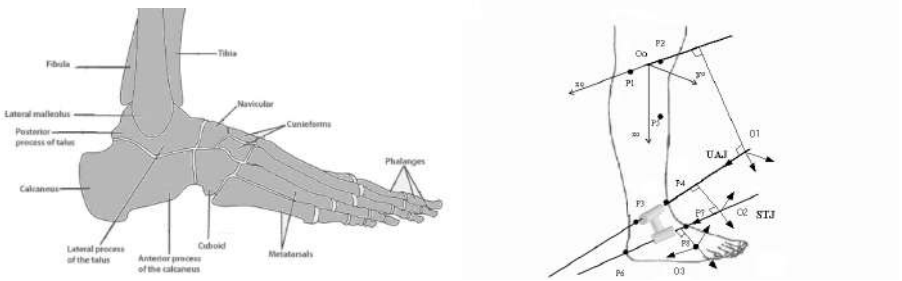
*Dissertation advisor: Ioannis Z. Emiris, Professor.

models, e.g. spherical joint and concludes with recent work e.g 4-bar modeling. In particular, certain advanced models incorporated the early findings of the two anatomical joint axes in the ankle, namely the upper ankle and the subtalar joint axis. In [8] was applied a 2-DOF (Degree Of Freedom) model to the lower leg, namely as a linkage of two ideal revolute joints and 3 rigid segments: the shank, the talus and the foot. The orientations axes could be determined by anatomical landmarks on the bones. In [1] were identified two subject-specific, fixed joint axes in the ankle complex by applying least-square optimization to minimize the difference between the real motion of external skin markers and the modeled motion. They noticed the relatively high variability between subjects especially of the subtalar and upper ankle joint orientation. Recently, a model that does not imply the existence of two distinct axes of the ankle joint complex and forms an innovative category of its own, have formulated a 2-dimensional model of the upper ankle joint by representing it as a closed 4-bar linked chain.

In closer detail, the ankle joint shows a mobile axes with a predetermined (1-DOF) path during passive flexion, but when loads are applied (e.g. muscle contraction), the mobile axes can deviate from this passive path, while the subtalar joint plays a stabilizing role. In our work, this deformable modeling is not required: the two-axes model represents the main movements and the kinematics of the ankle joint sufficiently well, and is quite accurate for our purposes. This has been proved in [16], in which was studied the use of the 2-axes ankle model and the identification procedure presented in [1] for calculation and comparison of the two hinge axes of the ankle joint complex for non-weight-bearing, weight-bearing and walking ankle motions. It was found that the 2-axes model fits the experimental data well with non-weight-bearing motion achieving the best fit. Since physiotherapy exercises especially in the early steps contain non-weighting-bearing movements the 2-axes model is sufficient for ankle modeling.

There have been a number of robotic devices proposed for ankle physiotherapy. Important work has been carried out at Rutgers University [10] with the development of a haptic interface for human ankle rehabilitation. This haptic interface has been based on a 6-DOF Stewart platform that applies variable forces and virtual reality exercises on the patient's foot, including remote control operation. However the Stewart platform is redundant for this application, the actuators used are noisy, the controller is oversized and the cost of the device is consequently high. In addition, in the rehabilitation program, there is no reference as to what extent the special characteristics of each patient's foot can be considered. Also the work in [3] is based on the study of ankle functional anatomy, which is represented in an orientation image space. Three parallel tripod-type ankle rehabilitation mechanisms were proposed. These are three or four actuator platforms and therefore they are redundant. Also, the rotation of the moving platform is performed about a vertical pivot strut, which is not a desirable characteristic for foot movements. In [27] was proposed an ankle rehabilitation device based on a reconfigurable parallel robot with 4-DOF and two moving platforms. However, this platform is quite complex and heavy and as a result is rather difficult in construction and transfer. A 2-DOF redundantly actuated parallel mechanism for ankle rehabilitation was proposed in [19]. The proposed device allows plantar- dorsiflexion and inversion-eversion using actuation redundancy to eliminate singularity, and to improve the workspace dexterity. However, this device is over-actuated, which means that there are redundant actuators.

The next important problem studied here is to determine the kinematics parameters of the ankle joint complex. Ankle joint complex kinematics is similar with that of a 2R serial manipulator (R denotes a revolute joint)[2] and therefore we approach the identification problem as a serial manipulator calibration problem. Identification of the axes of rotation of the limb-joints, by tracking specific point-markers on the body member, has received significant attention. In [1] was used the 2-axes model and identified the twelve parameters of an ankle joint complex model which is based on transformation matrices. This method requires the assignment of at least 3 non-collinear point-markers on each limb segment and tracking by the use of optical motion analysis systems. In [11] was proposed a least squares method by minimizing specific cost functions formed by the vector differences of the points markers positions. One drawback of the method is that for certain configurations of point markers (e.g. the markers distributed on a plane that contains the rotation axis) leads to ill-conditioned problems. Also, in [9] was presented a least squares method for average center and axis of rotation estimation. This method does not perform well if there is significant radial displacement from the center of the axis of rotation and needs at least three non-planar markers in general. These methods rely on position tracking of multiple point-markers which are assigned on the body by the use of expensive optical motion analysis systems.



(a) Foot anatomy and bones.

(b) Ankle rotation axes and the assigned D-H frames.

Figure 1: The ankle joint complex anatomy and kinematics.

2 Design specifications for an ankle rehabilitation robot

In order to design a robotic physiotherapy device the first step is to define the necessary design framework. For this we begin with the kinematics modeling of the ankle joint complex and we complete the design requirements with experimental data.

2.1 The ankle joint complex kinematics

The human ankle has a complex multi-joint structure. The central bone is the talus. Its surrounding bones are the calcaneus, the navicular and the cuboid; they are responsible for the rotation of the ankle joint in 3D (fig. 1(a)). The upper part of the talus articulates with the shank segment through the tibia and fibula bones. This is the upper ankle joint (UAJ); it supports the rotational dorsiflexion-plantarflexion motion. The movements between the fore bones are strictly coupled. Motion of the foot wrt the talus is regarded as a rotation about the (fixed) subtalar axis (STJ); this supports inversion-eversion motion.

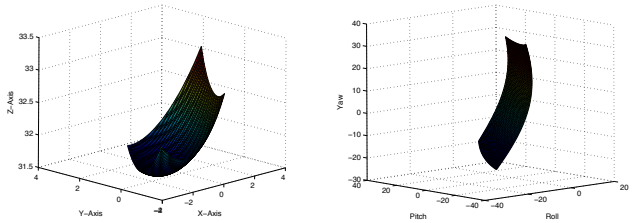
Among previous works, the early single-joint models are insufficient while the recent models do not fully describe foot rotations (e.g. they may consider only dorsiflexion-plantarflexion). Here we adopt the ankle joint complex model as a 2R serial manipulator [8], assuming the ankle rotation axes are straight lines through specific points. The lower limb is assumed to be composed of 3 rigid links capable to rotate between each other: the shank, the talus and the foot configuring a serial manipulator. The main movements of the foot are the plantarflexion-dorsiflexion and inversion-eversion. The size of foot bones and their relative positions as well as the orientation of rotation axes determine the foot kinematics. Many factors influence the joint rotation, e.g. shape of articular surfaces, position of rotation axes. Constraint and resistance on the foot motions are due to ligaments, capsules and tendons.

The parameters of this model are specified by a number of point markers that have been assigned on the human foot as in fig. 1(b). These point markers are used to obtain a set of distance measurements. We assign frame O_1 at the knee, centered between P_1, P_2, with the z-axis parallel to (P_1, P_2) and the x-axis vertical, passing through the midpoint of (P_3, P_4). By using the Denavit-Hartenberg(D-H) method [12, 22] we assign relative frames O_i between the moving links. T_i^{i+1} is the transformation matrix from O_i^{i+1} into O_i. The transformation matrix from the last into the first coordinate system is given from the relationship: $T_1^3 = T_1^2 T_2^3$. For a point $P = [x\ y\ z\ 1]^T$ on the last(foot) coordinate system the above transformation into the first(shank) coordinate system can be expressed as $P_o = [x_o\ y_o\ z_o\ 1]^T$:

$$P_o = T_1^3 P \tag{1}$$

from which the coordinates x_o, y_o, z_o are nonlinear functions $f_i(a_i, \alpha_i, d_i, \vartheta_i, x, y, z)$ of the D-H parameters $a_i, \alpha_i, d_i, \vartheta_i$. These equations give a parametric formula in the movement of P wrt the fixed coordinate system of the shank. The independent variables of the model are angles ϑ_2 (dorsiflexion-plantarflexion), ϑ_3 (inversion-eversion) while ϑ_1 is constant. According to the right-hand coordinate system assigned to the lower limb, the signs for rotation angles are: dorsiflexion(+), plantarflexion(-), eversion(+) and inversion(−). Movements of the left leg are assumed to be the mirror-image of the right leg [8]. The parameters α_i, a_i, d_i depend on the foot anatomy and size.

The transformation matrices were estimated for a male subject, and distances between the bony landmarks taken from [8]. From this data, a kinematics model of the foot was based on homogenous matrix transformations in Euler angles. Using calculations on the distances, we



(a) Translation *(b)* Rotation

Figure 2: Workspace from a point on the sole under the ankle.

obtain the D-H parameters. From the above analysis, the knee axis have angle 38.23^o with the UAJ and the UAJ with the STJ have angle 60.21^o.

We take ϑ_{dp}, ϑ_{ie} as the new variables for the dorsiflexion-plantarflexion and inversion-eversion angles measured from the standing posture, and insert $\vartheta_2 = \vartheta_2^o + \vartheta_{dp}$, $\vartheta_3 = \vartheta_3^o + \vartheta_{ie}$ into eqts (1). Now ϑ_{dp}, ϑ_{ie} are the input variables of the model. Common ranges for the movement are $-40^o \leq \vartheta_{dp} \leq 30^o$, $-20^o \leq \vartheta_{ie} \leq 20^o$ [22]. Based on the model, we specify the foot workspace when inputs range through all possible motions. Our first requirement is the shank to be fixed and vertical wrt the World Coordinate system attached to the base of the robot.

We fix point P_f on the sole under the ankle where the center of the moving platform will be attached. We assume P_f is on the positive axis of the knee's frame and has a distance equal to this of P_6. The workspace produced by the foot will be derived from the motion study of P_f. By eqts (1) and letting inputs $\vartheta_{dp}, \vartheta_{ie}$ run through their entire regions, P_f traces the surface of fig. 2(a). Feet of every size and anatomy produce the surface in fig. 2(a). The geometric characteristics of this surface (e.g. shape, curvature), depend on α_i, a_i, d_i. Every trajectory traced by P_f is within this surface.

Table 1: Coordinate ranges of a point on the sole under the ankle.

ΔX=56 mm		ΔY=41.7 mm		ΔZ=17.3 mm	
Min X	Max X	Min Y	Max Y	Min Z	Max Z
-33.5	22.4	-19.6	22	-2.8	14.4

We compute the orientation of the foot when its axes are rotated in specific angles. First, we establish a reference frame with its origin at P_f. The axes are parallel with those of the base frame when the foot is in the neutral position. The rotation angles roll(α), pitch(β), yaw(γ) of this frame wrt the base frame are the rotation angles of the moving platform. The foot model, when ϑ_{dp}, ϑ_{ie} take all values in their ranges, yield the rotation workspace in fig. 2(b). By

Table 2: Orientation ranges based on the model.

$\Delta\alpha$=30.56 deg		$\Delta\beta$=76.58 deg		$\Delta\gamma$=62.49 deg	
Min α	Max α	Min β	Max β	Min γ	Max γ
-20.71	9.84	-39.95	36.63	-25.34	37.15

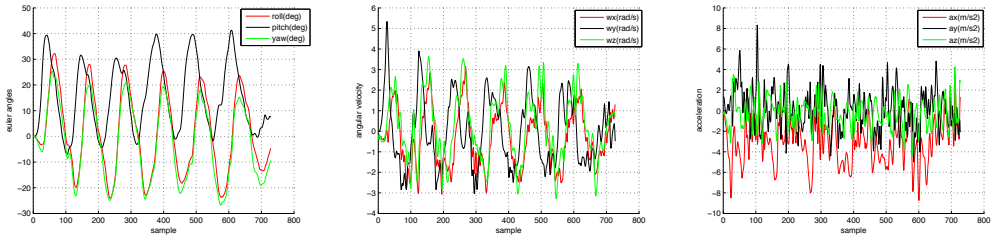
assuming that the angle axes parameters in eqts (1) are found in well specified intervals, we will specify the extended workspace produced by the model (eq. (1)). In [14], the orientation of the lower limp rotation axes, and the ranges in the relevant angles between them are measured. The results depend on the position of the foot even for a given patient. Different patients will give different results. We conclude that the model parameters are quite uncertain and so the model must be extended to include uncertainties. By computing the minimum and maximum values we take the values in table 3.

2.2 Experimental data

A Mephisto 3D Scanner was used to take images of the right foot sole of 11 adult healthy human subjects of different age, height, weight and gender. We used 5 positions: Neutral, Right-Up, Right-Down, Left-Up and Left-Down. The reference is a central point on the sole under the ankle because this point will be controlled by the platform. The coordinate differences among

Table 3: Extended coordinate ranges of a point on the sole under the ankle.

ΔX=114.2 mm		ΔY=98.9 mm		ΔZ=37.3 mm	
Min X	Max X	Min Y	Max Y	Min Z	Max Z
-67.6	46.9	-4.55	53.4	-7.6	29.7



(a) Euler angles *(b)* Angular velocities *(c)* Linear accelerations

Figure 3: Recorded data in full rotations of the foot.

the point's positions are found inside the following enclosing volume: ΔX=12cm , ΔY=12 cm, ΔZ=12 cm.

To measure orientation angles, rotation velocities, and accelerations, we performed experiments with an MTi motion sensor of XSens Motion Technologies. It provides and records pitch, roll and yaw angles, rate of turn and linear accelerations in axes X, Y, Z. We used the right foot of 5 adult healthy humans of both genders and different heights, with the shank kept vertical and fixed. The only moving part was the foot. The sensor was attached on the sole of the foot under the ankle. Data were recorded during dorsiflexion-plantarflexion and inversion-eversion throughout the entire range of movement. Fig. 3(a) shows roll, pitch and yaw wrt time of a foot, in extreme rotational movement. Fig. 3(b) shows angular velocities wrt time in extreme rotational movement while fig. 3(c) shows linear accelerations wrt time in extreme rotational movement. The maximum recorded angular velocity was 9.3rad/sec and therefore the upper bound of 10 rad/sec is adopted.

The torque bounds are coming from the literature in which they studied the tension torque that the soleus and tibialis anterior muscles can exert. These are two of the main dorsiflexor-plantaflexor muscles and the maximum measured torque was about 121Nm. Also, in another work the maximum measured torques of the whole plantarflexor and dorsiflexor muscle groups were about 143Nm [22, 24]. Therefore a desired upper bound of 200Nm includes a wide range of foot torque capabilities. Thus, our platform will operate up to 200Nm, to handle torque-producing tasks at different velocities during concentric or eccentric muscle actions.

3 Design of a 2-DOFs hybrid parallel-serial ankle physiotherapy robot

Initially, we studied two existing robotic architectures as possible ankle physiotherapy devices [23]. The first one was the Agile Eye. Although the Agile Eye has only 3 rotational DOFs and large workspace, its sensitivity to transfer and its rotations about only one point led us to reject this robot. The second one was a parallel Tripod(3-RPS) with an extra rotation axis on the moving platform as a possible ankle rehabilitation device was studied in . The Tripod has two rotational (pitch, roll) and one translational (z) degrees of freedom. As the yaw angle changes significantly during the foot movements on the platform, an extra rotation axis was added on the moving platform to provide the necessary extra yaw angle. Although this device can not follow the foot movements satisfactorily or include mechanical adjustability and our effort is to design a robot with fewer DOFs.

To overcome the previous disadvantages we proposed a hybrid serial-parallel robotic architecture with 2-DOF and mechanical adjustability [24]. The robot consists of a base platform and a moving platform like most parallel robots. The latter is where the patient's foot shall be placed. A vertical strut connects the base of the robot with a passive serial chain. The serial chain has structure similar to that of the foot and provides the necessary constraints on the movements. It has one revolute and one cylindrical joint which support the rotations about the two main rotation axes of the ankle, see Figure 4. R_1 is a revolute-joint which is collinear with

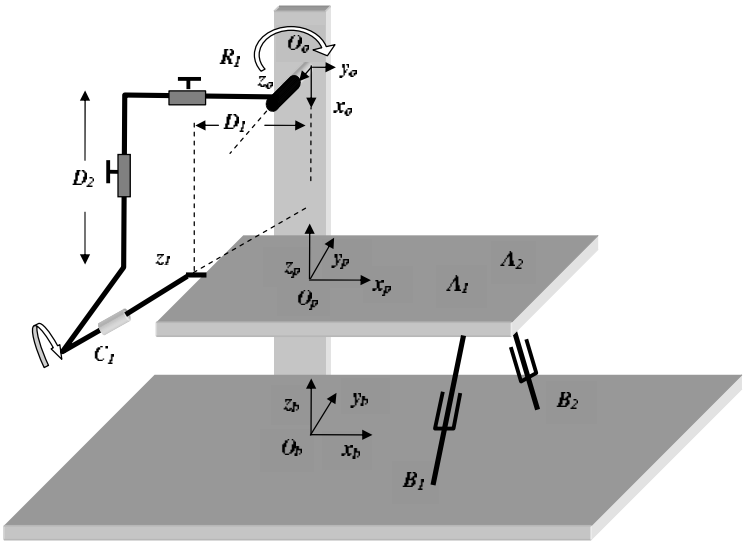


Figure 4: The hybrid serial-parallel 2DOF ankle physiotherapy robot.

the Upper Ankle Joint (UAJ) and C_1 is a cylindrical-joint which is collinear with the Subtalar Joint (STJ) of the foot. The serial chain is connected with the moving platform and has two adjustable screws so that the corresponding lengths D_1, D_2 can be adjusted according to the axes position of each individual patient's foot. The parallel chain consists of two prismatic actuators which are connected with the moving platform through S-Joints (points A_1, A_2) and with the base platform through U-Joints (points B_1, B_2).

3.1 Kinematics modeling of the platform

We start by modeling the kinematics of our device. The mobility of the platform is modeled by applying the Grubler formula for spatial structures. The total number of its degrees of freedom N is given as follows: $N = 6(n-j-1) + \sum_i^n f_i = 6(8-9-1) + 14 = 2$, where n represents the total number of rigid bodies of the mechanism, including the base, j is the total number of joints, and f_i the number of degrees of freedom of joint i. Initially we assign the base coordinate frame $O_b x_b y_b z_b$ on the fixed base and the moving frame $O_p x_p y_p z_p$ on the moving platform as shown in figure 4. The two frames are parallel when the moving platform is in zero position. As we have a serial kinematics chain it is useful to implement the Denavit-Hartenberg method [12] for the assignment of the relative reference frames on the passive serial chain and therefore to obtain the overall kinematics formula of the platform. The $O_o x_o y_o z_o$ frame is the base frame of the serial chain and is placed arbitrarily on the strut with the axis z_o collinear with the UAJ axis, and x_o collinear with z_b, as shown in Figure 4. The origin O_2 of the $O_2 x_2 y_2 z_2$ frame is the O_p of the platform frame and the axis z_2 is parallel with z_1 (C_1). The total transformation matrix T is given by the multiplication: $T = T_b T_0^1 T_1^2 T_r$, where T_b is a constant homogeneous rotation matrix defining the relative rotation of the $O_o x_o y_o z_o$ frame into the base $Ob_x b_y b_z b$ frame. The points on the platform coordinate frame need to be multiplied with an extra rotation matrix T_r in order to be transformed into the last D-H frame of the serial chain. The inverse kinematics problem is described from the following two equations:

$$L_i^2 = \|B_i A_i\|^2 = \|T \cdot A_i - B_i\|^2, \tag{2}$$

where i=1,2 and L_i, are the length of the actuated links. It is well known that the Jacobian matrix is a key part for the study and design of robots. By following the normalized Plucker vector-based procedure for the inverse Jacobian calculation of a general parallel robot as it is described in detail in [18] we get:

$$J^{-1} = \begin{bmatrix} n_1 & n_1 \times A_1 O_p \\ n_2 & n_2 \times A_2 O_p \end{bmatrix}, \tag{3}$$

where n_1, n_2 are the unit vectors of $B_1 A_1$, $B_2 A_2$.

3.2 Velocity and Force Transmission

When the linear actuators are activated their velocities and applying forces are transferred onto the moving platform. The relations between the forces-velocities of the actuators and

the moving platform are expressed by the Jacobian matrix and are pose-dependent. Here we follow the kinetostatic capability analysis of parallel manipulators described in [15] where the calculation of the magnitude bounds of the force and velocities of the end-effector is reduced to an eigenvalue problem. Matrix M is essential here and is defined as follows:

$$M = J^{-T}J^{-1} = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \tag{4}$$

Given the desired velocities as well as the force and moments that should be handled by the platform, the extreme values of the velocities and forces of the linear actuators are computed in the design phase. The linear forces and torques have different units it is reasonable for the bounds computation of force-torques at the end-effector to be decoupled into two constraint maximization subproblems, one for the forces f and one for the torques m. By use of matrix M the two maximization subproblems can be rearranged into eigenvalue subproblems. Following a similar procedure as in the force transmission analysis, velocity transmission analysis is decoupled in two subproblems, one for the linear ν and one for the angular ω velocities magnitudes.

$$\frac{\|f\|}{a_{fmax}} \leq \|\tau\| \leq \frac{\|f\|}{a_{fmin}}, \quad \frac{\|m\|}{a_{mmax}} \leq \|\tau\| \leq \frac{\|m\|}{a_{mmin}} \tag{5}$$

$$a_{\nu min}\|\nu\| \leq \|\dot{L}\| \leq a_{\nu max}\|\nu\|, \quad a_{\omega min}\|\nu\| \leq \|\dot{L}\| \leq a_{\omega max}\|\nu\| \tag{6}$$

where A, C are the 3×3 submatrices of M and ν, ω the 3×1 vectors of linear and angular velocities of the end-effector, f, m the 3×1 vectors of linear forces and torques on the end-effector. The bounds of the velocities magnitudes are therefore given from the inequalities where $a\nu_{max}$ and $a\nu_{min}$ denote the square roots of the maximum and minimum eigenvalues of A, and $a\omega_{max}$ and $a\omega_{min}$ are those of the maximum and minimum eigenvalues of C. By discretizing the whole workspace of the robot and computing the global extreme eigenvalues of M, the global magnitude bounds of velocities and forces are computed.

3.3 Parametric Design of the Robot

This section presents the parametric design of our mechanism, since we have chosen the robotic architecture. Design concerns the calculation of the geometric parameters of the robot satisfying our requirements [18]. Based on the above foot analysis, the following values were selected for initial dimensioning of the device: Moving platform: 0.40x0.20m so that it can accept all or at least the majority of human foot sizes. Base platform: 0.60x0.40m. The first axis zo is placed 50cm above the fixed base. The bounds of rotation axes of the serial chain were defined according to the range of the feet rotations. The STJ axis, and so the z1 has the mean foot orientation as it is given and forms an angle of 23^o with the x_pz_p plane and an angle of 41^o with the x_py_p plane. This work has been extended and in order to complete the design of the robot, additional measurements on the foot of several human subjects have been conducted. Coordinates of specific points of the foot, have been measured, utilizing a Microscribe coordinate measuring device. For the experimental measurements the right feet of 19 adult males and females have been used in the erect standing pose. The UAJ is defined by points P_3 (lateral malleolus) and P_4 (medial malleolus), while the STJ is defined by points P_6 (calcaneus point) and P_7 (navicular point). For calculating the bounds on distance D_1 the points P_3, P_4 and P_6 were projected on the horizontal plane. The vertical distance of the projected point P_6 from the projected line P_3P_4 defines distance D_1. The computed values of D_1 were found to be in the range: $3.5cm \leq D_1 \leq 5.6cm$, with mean value 4.83cm and standard deviation 0.68cm. For calculating the range of distance D_2, the mean value of the height of points P_3 and P_4 from the horizontal plane was computed. The resulting values are in the following range: $5.4cm \leq D_2 \leq 9.3cm$, with mean value 7.29m and standard deviation 1.02cm. Points A_1, A_2 on the moving platform have been assigned coordinates (0.15,-0.06,0) and (0.15,0.06,0) of the moving frame respectively. If the points are far from the rotation axis then the actuators apply smaller forces but larger velocities. Reversely, if the points are near the rotation axis then the actuators apply forces with larger values and smaller velocities. Therefore, the points are selected to be in the middle of the platform in order to balance the amounts of the velocities and forces exerted by the actuators. The coordinates of the base platform points B_1, B_2 have been assigned to (0.10,-0.15,0) and (0.10,0.15,0) respectively, on the base reference frame nearer to the origin so to avoid singularities. The coordinate units are in m. When the platform moves through the entire range of rotations and parameters D_1, D_2 take all values in the above intervals, then the length of the legs are found in the range: [0.31m, 0.56m]. Having computed the kinematics

parameters of the robot as well as the desired end-effector velocities and wrench forces, the actuator velocities and forces can be computed. The platform must handle torque values up to 200Nm. In order for the platform to achieve these torque bounds, the actuator forces must be greater than 675N according to equation (5). Similarly, for velocities calculation used the upper bounds of 10rad/sec in angular velocities of the platform. According to equation (6), the linear actuators should achieve velocities at least 2.1m/sec in order characteristics providing the robot motion can be selected. Simulations with the maximum design values movements shows that the platform fulfils the design requirements, proving the succeed of design [24].

4 Identification of the ankle joint complex kinematics

For the identification of the ankle joint complex kinematics parameters it is useful to implement robot calibration techniques. Since the internal joint ankle values is difficult to be measured, conventional calibration techniques are not applicable. Therefore, we approach the problem by trying algebraic elimination and trajectory fitting in 3D circles [21].

4.1 Algebraic methods

We first approached identification via algebraic variable elimination. Algebraic elimination of the two rotation angles of the ankle might be a reliable approach, just as it has been for parallel robots [4, 5]. However, the case of serial robots presents certain difficulties compared to parallel robots. In the latter case, the kinematics equations are produced by one multiplication with one homogeneous transformation matrix. In serial robots, the kinematics equations are obtained by successive multiplication with as many homogeneous matrices as the number of links. Hence, the final kinematics equations are quite complex [21].

One approach is to linearize the polynomial system by resultants. It is typically expressed as a matrix determinant, and the resultant matrix can be used to reduce system solving to a problem in linear algebra [6]. In the case of a manipulator with two revolute joints, its kinematics model is the product of two D-H matrices. To eliminate the two rotation angles by resultants, the *mresultant* function of MAPLE package *Multires* yields a 21×21 matrix which contains polynomials of degree up to 15, therefore the determinant is very hard to compute.

4.2 Identification by nonlinear fitting in 3D circles

This section presents identification without using internal joint values, based on fitting 3D circles. Let us consider a serial manipulator with only revolute joints, where all the joints are fixed in their zero position. Starting from the last axis we rotate it and record the Cartesian position of the end-effector. After we fix the last axis, we rotate the previous axis while all the remaining axes are unmovable in the zero position and record the end-effector's positions. We repeat until the rotation of the end-effector about the first axis is recorded. Rotation of a point about an arbitrary axis in 3D traces a circular arc. Given N measured points of the arc, we estimate the parameters (center, radius, normal) in two steps: (1) the points are fitted on the plane of the arc, while the plane-normal defines the rotation axis, (2) we compute the center and radius of the arc.

Nonlinear least-squares minimize $f = \sum_{i=1}^{N} e_i^2$, where the e_i denote the errors, by iteratively linearizing around the parameters the following:

$$\Delta f = J_k \Delta x, \tag{7}$$

where Δf is the error between measured and residual function, Δx is the correction of parameter vector x in the current estimate, and J_k is the identification Jacobian. Usually, weighting of least-squares gives better accuracy and is achieved by task variable and column scaling. The measured data are of the same units and therefore task variable scaling via noise covariance matrices is not necessary. On the other hand, column scaling improves the condition number $\kappa = \sigma_1/\sigma_p$, where σ_1 is the largest and σ_p the smallest nonzero singular values of J_k. Column scaling does not affect the solution and is achieved by right multiplication of J_k by matrix $H = diag(h_1, \ldots, h_n)$ with

$$h_i = \begin{cases} ||J_{ki}^{-1}||, & \text{if } ||J_{ki}|| \neq 0, \\ 1, & \text{if } ||J_{ki}|| = 0\ , \end{cases} \tag{8}$$

where J_{ki} is the i-th column of J_k. With column scaling the singular values of J_k become comparable. Studying J_k before the actual identification, provides information about identifiability and the observability.

a) Identifiability: This determines whether all parameters can be identified independently, or some are non-identifiable because depend on others. Non-identifiable parameters exist if J_k is rank deficient. In this case, several approaches have been proposed [13] for identification: (i) elimination of the non-identifiable parameters by QR decomposition (ii) zeroing small singular values, and (iii) incorporating a priori estimates, where small or zero singular values are counteracted by a constant weighting factor. An efficient technique we use here is (ii), which identifies without removing any parameter or a priori estimates. Our method begins with singular value decomposition (SVD) of scaled matrix $J_k = U\Sigma V^T$, where U is $N \times N$ orthogonal, V is $n \times n$ orthogonal, and Σ is $N \times n$ diagonal, with singular values $\sigma_1 \geq \cdots \geq \sigma_p > 0 = \cdots = 0$. If $p < n$, then J_k is rank deficient so each iteration step of 7 computes

$$\Delta x = \sum_{j=1}^{n} \frac{u_j^T \Delta f \, v_j}{\sigma_j}, \tag{9}$$

where, if σ_j is zero or numerically small, we set $1/\sigma_j = 0$.

b) Observability: Observability concerns the selection of the best measurement sets according to an observability index. Several observability indices which indicate the measurements of high accuracy have been proposed. A very interesting one is the *noise amplification index* $O = \sigma_p^2/\sigma_1$, which relates the amplification of the measurement noise with the estimated parameters. It is more sensitive to measurement and modeling error than previous indices. The measurement poses set with larger O results to a more accurate identification.

Step 1: The coordinate data will be fitted to the 4 parameter plane: Ax+By+Cz+D=0. The vertical distance d_i of a 3D point i from a plane is given by:

$$d_i = \frac{Ax_i + By_i + Cz_i + D}{\sqrt{A^2 + B^2 + C^2}}. \tag{10}$$

We have to minimize the objective: $f_d = \sum_1^N d_i^2$. Identification is performed after certain iterations of linear least-squares 7. The identification Jacobian is the stacked matrix for all measurements of the rows J_k^l of the following derivatives : $J_k^l = \left[\begin{array}{cccc} \frac{\partial d_i}{\partial A} & \frac{\partial d_i}{\partial B} & \frac{\partial d_i}{\partial C} & \frac{\partial d_i}{\partial D} \end{array} \right]$.

Step 2: The arc points are projected onto the plane by computing a verticality condition. The center and radius of the arc is defined by an intersecting sphere with the resulting plane of *Step 1*. The residuals of the sphere and measured data points are: $e_i = (x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2 - R^2$. Therefore, to find the best fitting sphere, we solve the constrained nonlinear least-squares: $f_e = \sum_1^N e_i^2$, subject to: $h(x_c, y_c, z_c) = Ax_c + By_c + Cz_c + D = 0$. Similarly with the linearized procedure of *Step 1*, the constrained nonlinear least-squares can be solved by iteratively solving:

$$\begin{bmatrix} -J_s^T J_s & J_h^T \\ J_h & 0 \end{bmatrix} \cdot \begin{bmatrix} p_k \\ \lambda_k \end{bmatrix} = \begin{bmatrix} J_s^T e_i \\ -h \end{bmatrix}, \tag{11}$$

where J_s the identification Jacobian, namely the stacked matrix for all measurements of the rows J_s^l of the following derivatives of the residuals:$J_s^l = \left[\begin{array}{cccc} \frac{\partial e_i}{\partial x_c} & \frac{\partial e_i}{\partial y_c} & \frac{\partial e_i}{\partial z_c} & \frac{\partial e_i}{\partial R} \end{array} \right]$, where matrix J_h is the Jacobian of the constrained equation h with respect to the parameters (x_c, y_c, z_c), $p_k = [x_c, y_c, z_c, R]$ is the vector of estimated parameters, and λ_k is the scalar Lagrange multiplier for the constraint. The two steps yield the circle center $p_c = (x_c, y_c, z_c)$, radius R, as well as the normal vector $\overrightarrow{n}$, which is the unit vector of $\overrightarrow{p} = [A, B, C]$.

4.3 Identification by linear fitting in 3D circles

We follow the previous two-steps procedure of fitting the plane and circle but avoid to use non-linear least-squares.We employ direct methods which are computationally faster and accurate.

Step 1: The plane is estimated via SVD [20]. Let p_p be a point on the best fitted plane, p_i a given point, and $\vec{a}$ the normal vector to the plane. The vertical distance of p_i from this plane is $d_i = \vec{a} \cdot (\vec{p_i} - \vec{p_p})$, so minimizing $J = \sum_1^N d_i$ yields $\vec{a}$ subject to $|\vec{a}| = 1$. By Lagrange multipliers, one obtains a 3×3 eigenproblem: $(M^T M)\vec{a} = \lambda \vec{a}$, where $M = [x_i^T y_i^T z_i^T]$ is the $N \times 3$ data matrix. Eventually, we get: $\sum_i (\vec{a} \cdot \vec{p_i}) = \lambda |\vec{a}|^2 = \lambda$, so minimizing J is reduced to computing the minimum eigenvalue of $M^T M$. This is the square of the minimum singular value of M, computed by SVD, which is quite stable numerically. The corresponding singular vector is $\vec{a}$. Since the centroid of the data belongs to the plane, we can specify all parameters of the plane equation.

Step 2: The center and radius of rotation will be estimated by applying 2D methods. After the estimation of the best plane, the measured data are projected onto the plane. A coordinate frame is defined as follows: $\overrightarrow{a}$ is the unit vector through p_1, p_n, $\overrightarrow{n}$ is the unit normal vector of the plane, $\overrightarrow{o} = \overrightarrow{n} \times \overrightarrow{a}$. The following matrix maps the arc points from the frame of $\overrightarrow{a}, \overrightarrow{o}, \overrightarrow{n}$, to the base frame of the sensor:

$$T = \begin{bmatrix} \overrightarrow{a} & \overrightarrow{o} & \overrightarrow{n} & \overrightarrow{p_1} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

Now, we have to estimate the circle in 2D. After estimation of the center and radius, the former is mapped through T back to 3D space. Here we compare existing nonlinear and linear methods for 2D arc fitting. These methods are: (1) Algebraic circle fit with "hyperaccuracy", (2) Karimaki's,(3) Landau's,(4) Pratt's with SVD,(5) Riemann sphere fitting, (6) Levenberg-Marquard, (7) the Trust region,(8) Linear least squares by Kasa, and (9) Thomas' method.

The previous methods are evaluated, on a 2D arc estimation. An arc of 60^o is selected and normal distributed noise with standard deviation $\sigma = 0.06$ was added. As evaluation factors for the methods we selected (a) the distance between the estimated center and the actual center, (b) the difference of the estimated radius from the actual one, and (c) execution time. The Riemann sphere fitting method was more accurate and fast, because it is non-iterative and has been selected for arc fitting.

The fitting with Riemann-spheres [17], is based on the 1-to-1 mapping of a 2D circle to a Riemann sphere through stereographic projection. A circle on the Riemann sphere is its intersection with a plane in 3D, therefore the fitting of a 2D circle is reduced to fitting a plane in the Riemann sphere. The procedure is non-iterative, fast, accurate, and has 3 steps [21].

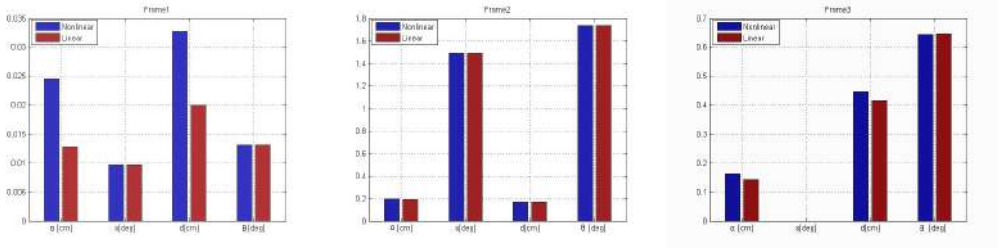


Figure 5: Simulation evaluation of the linear and nonlinear methods with noise ($\sigma = 0.04$).

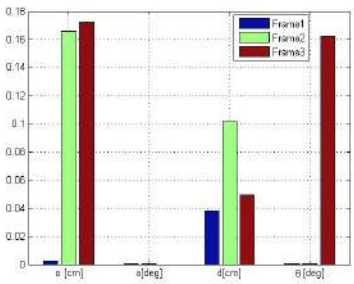


Figure 6: Experimental evaluation of the linear and nonlinear identification.

Both nonlinear and linear methods yields the equations of the 2 ankle axes. The D-H parameters can be computed geometrically based on the definition of the method [21]. **Results:** Initially the proposed method is evaluated on the model. First, the inversion-eversion motion were executed taking all the values in its range producing 41 points. Next, we fix STJ, and UAJ is rotated in its full range, producing 66 dorsiflexion/plantarflexion data of point P_8. Normally distributed noise with standard deviation ($\sigma = 0.04$) was added on the measured data and the simulation results are shown in fig 5. For the plane estimation of *Step 1* the observability index O was evaluated on several position sets. The position set with many equally distributed points along the arc gave the maximum O value. By performing SVD on the identification Jacobian we found that only the 4th singular value vanishes. Similarly, for the center estimation of *Step 2* several position sets were tested. Position set with many equally distributed points along the arc gave the maximum O value. The identification Jacobian in *Step 2* has full rank and there were no non-identifiable parameters.

Our method is evaluated on an actual human foot. An adult male subject of height 1.85m and weight 80kg has been used, and the motions of the right foot were measured. For recording, we used a Microscribe 5-axes coordinate measurement device. For the dorsiflexion-plantarflexion motion, 13 positions were recorded, while for the inversion-eversion motion, we recorded 12. In comparison with the simulation results (fig. 6), the deviation between the two methods is larger which emanates from the smaller number of measurement points.

5 Conclusion and future work

In this thesis was studied the design and identification of kinematics parameters of robotic mechanisms for ankle physiotherapy applications. Initially, we adopted the suitable 2 axes ankle joint kinematics model from the literature and we received experimental data, in order to define a design framework for the design of a rehabilitation robot. Then, we introduced a new 2-DOF hybrid serial-parallel robot with mechanical adjustability and we performed its parametric design. This physiotherapy device outperforms to the existing physiotherapy robots on the minimum number of actuators, safety, modularity and economy. Finally, we developed a simple, robust and accurate method for identification of the ankle joint complex kinematics parameters for the appropriate tuning of a physiotherapy robot. Our method avoids the use of position tracking of multiple point-markers on the body-member and the use of expensive optical motion analysis systems. This makes it applicable in a physiotherapy clinic.

As future work, we may consider the structural study and construction of the robot. We have to investigate appropriate compliance control laws (e.g. stiffness, impedance, hybrid position/force control) [25, 26] in order to evaluate their suitability for physiotherapy operations. The development of the user software with teleoperation capabilities and operation of the platform with patients under the supervision of physiotherapist are in our interests. Also, we may revisit algebraic techniques, investigating sophisticated algorithms such as sparse interpolation, or parameter reduction so as to reduce the complexity of the resultant calculations in the kinematics parameters identification of the ankle joint complex.

Acknowledgements

We thank Dr. C.N. Maganaris of IRM Manchester Metropolitan University for his useful advice. This work was supported by the General Secretariat of Research and Technology of Greece through a PENED 2003 program, contract nr. 70/03/8473, co-funded by the European Social Fund (75%), national resources (25%).

References

[1] Van Den Bogert AJ, Smith GD, Nigg BM (1994). In vivo determination of the anatomical axes of the ankle joint complex: An optimization approach. J Biomech, 27:1477–1488.

[2] Craig J (1989). *Introduction to Robotics: Mechanics and Control.* Addison Wesley.

[3] Dai JS, Zhao T, Nester C (2004). *Sprained Ankle Physiotherapy Based Mechanism Synthesis and Stiffness Analysis of a Robotic Rehabilitation Device.* Autonomous Robots, (16)207-218.

[4] Daney D, Emiris IZ (2001). Robust parallel robot calibration with partial information. In Proc IEEE Intern Conf Robotics Automation, Seoul, S. Korea, pp. 3262–3267.

[5] Daney D, Emiris IZ (2004). Algebraic elimination for parallel robot calibration. In *Proc IFToMM Congress Mechanism & Mach Science*, Tianjin, China.

[6] Emiris IZ (1994). Sparse Elimination and Applications in Kinematics. PhD Thesis, Computer Science Division, Univ. of California at Berkley.

[7] Dettwyler M, Stacoff A, Kramers-de Quervaina IA, Stussi E (2004). Modelling of the ankle joint complex. Reflections with regards to ankle prostheses. Foot & Ankle Surgery., 10:109–119.

[8] Dul J, Johnson GE (1985). A kinematic model of the human ankle. J Biomedical Engineering, 7:137–143.

[9] Gamage SSHU, Lasenby J (2002). New least squares solution for estimating the average center of rotation and the axis of rotation. J Biomech 2002, 35:87–93.

[10] Girone M, Burdea G, Bouzit M, Popescu V, Deutsch JE (2001). A Stewart platform-based system for ankle telerehabilitation. Autonomous Robots, 10:203–212.

[11] Halvorsen K, Lesser M, Lundberg A (1999). A new method for estimating the axis of rotation and the center of rotation. J Biomech, 32:1221–1227.

[12] Hartenberg R, Denavit J (1955). A kinematic notation for lower-pair mechanisms based on matrices. J Appl Mech., 22:215–221.

[13] Hollerbach JM, Wampler CW (1996). The calibration index and taxonomy for robot kinematic calibration methods. Int J Robotics Research, 15:573–591.

[14] Isman RE, Inman VT (1969). Anthropometric Studies of the Human Foot and Ankle. Bulletin of Prosthetic Research, 10-11:97–129.

[15] Kim HS and Choi YJ (1999). The Kinetostatic Capability Analysis of Robotic Manipulators. in Proc. IEEE/RSJ Int. Conf. Intelligent Robots & Systems., 55-67, Korea

[16] Leitch J, Stebbins J, Zavatsky AB (2010). Subject-specific axes of the ankle joint complex. J. Biomechanics., 43:2923–2928.

[17] Lillekjendlie B (1997). Circular arcs fitted on a Riemann Sphere. Computer Vision & Image Understand., 67:311–317.

[18] Merlet J-P (2006). Parallel Robots. Springer Second Eds.

[19] Saglia J, Tsagarakis NG, Dai JS, Caldwell DG (2009). A High Performance 2-dof Over-Actuated Parallel Mechanism for Ankle Rehabilitation. Proc. IEEE Int. Conf. Robotics & Automation, Kobe, Japan, 805-809.

[20] Shakarji CM (1998). Least-squares fitting algorithms of the NIST algorithm testing system. J Res Natl Inst Stand Technol, 103:633–640.

[21] Syrseloudis CE, Daney D, Emiris IZ (2011). Identification of Kinematics Parameters of the Ankle Joint Complex. Submitted for Journal Publication

[22] Syrseloudis CE, Emiris IZ, Maganaris C, Lilas T (2008). Design framework for a simple robotic ankle evaluation and rehabilitation device. In Proc IEEE Intern Conf Engineering in Medicine & Biology, Vancouver, 4310–4313.

[23] Syrseloudis CE, Emiris IZ (2008). A Parallel Robot for Ankle Rehabilitation-Evaluation and its Design Specifications. In 8th Int. Conf. IEEE in Bioinformatics and Bioengineering, Athens, 106.

[24] Syrseloudis CE, Emiris IZ, Lilas T, Maglara A (2011). Design of a simple and modular 2-DOF ankle physiotherapy device relying on a hybrid serial-parallel robotic architecture. J Appl Bionics & Biomech, Special issue, 8:1-14.

[25] Tzafestas SG, Syrseloudis CE, Rigatos GG (2000). Hybrid Position Force Control via a Neuro-fuzzy Technique: Application to the Milling Process. *Intern. J. Mechanical Production Systems Engineering*, No 4, pp. 25-36.

[26] Tzafestas SG, Syrseloudis CE, Rigatos GG (2000). Neuro-fuzzy Hybrid Position/Force Control of Industrial Robots: A Simulation Study for the Milling Task. *3d IMACS Symp. Mathematical Modeling*, Vienna, Austria, February.

[27] Yoon J, Ryu J (2005). A Novel Reconfigurable Ankle/Foot Rehabilitation Robot. IEEE Int. Conf. Robotics & Automation. Barcelona, Spain, 2290-2295.

Web-based Adaptive Learning Environments and Open Learner Model – Use in Didactics of Informatics

Ilias Verginis *

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
iliasver@di.uoa.gr

Abstract. In the context of this dissertation, issues concerning adaptive Web-based learning environments, the Open Learner Model and its exploitation towards the enhancement of the teaching and learning process are studied. The thesis focuses on the research of Open Learner Models in adaptive Web-based learning environments, aiming to design develop and exploit the open learner model of the Web-based, learning environment SCALE (OLM_SCALE). In the frame of this thesis two empirical studies were conducted: The first one aims to investigate how the educational material in form of activities that exploit the learning design of SCALE, can support the learning process in the context of the undergraduate course "Introduction to Informatics and Telecommunication". The second one aims to investigate how the facilities offered by OLM_SCALE can be exploited towards the reengagement of disengaged students.

Keywords: Open Learner Model, Web-based Learning Environments, Introductory Computer Science Courses, Learning Activities, Tutoring Feedback Components.

1 Introduction

The contemporary tendencies for supporting and promoting students' learning process in undergraduate curricula suggest the use of learning environments [1], [2], [3], [4]. A negative aspect of this trend is that students might become disengaged when using tutoring software and try to *game* the system by moving rapidly through problems without really studying them and by seeking the final hint that might give the answer away [5]. Recognizing this fact many researchers have placed focus on developing pedagogical approaches for the detection and guidance of online students who become disengaged. The majority of those approaches are based on models that are trained through extensive analysis of the log files that represent students' interaction with the learning environment. Specifically Cocea and Weibelzahl in [6] propose the exploitation of several data mining techniques in order to find the best method and the indicators for disengagement prediction. The authors argue that motivational level could be predicted from very basic data commonly recorded in log files, such as events related to reading pages and taking tests. The students identified to be disengaged are engaged in a dialog in order to assess their self-efficacy, self-regulation and other related motivation concepts. Baker et. al. in [7] propose a machine-learned Latent Response Model that is highly successful at discerning which students frequently *game* the system in a way that is correlated with lo[1]w learning. Specifically the research team uses three data sources in order to train the model to

* Dissertation Advisor: Maria Grigoriadou, Professor

predict how frequently a student gamed the system. The results of the empirical study shows that the model is successful at recognizing students who game the system and show poor learning. Johns and Woolf in [8] propose a dynamic mixture model based on Item Response Theory (DMM-IRT) to detect students' motivational level and estimate their proficiency. The results of the corresponding experiments suggest that the DMM_IRT model can better predict students responses compared to a model that does not account for motivation. In the work of [9], a detailed log file analysis is used as input for the actions performed by the animated agent named "Scooter the Tutor". Scooter interacts with the student (by expressing negative emotion to gaming students), aiming to reduce the incentive to game, and help students learn the material that they were avoiding by gaming, while affecting non-gaming students as minimally as possible. Whenever "Scooter" detects a *gaming* student, he provides him/her with supplementary exercises focused on exactly the material the student bypassed by gaming.

Although the aforementioned approaches manage to identify and guide the disengaged students, they require time consuming and skilfully log file analysis in order to retrieve data suitable for training the specific models. Since a web based learning environment can generate thousand lines of information per hour, specific applications designed to analyze and impact meaning to raw log file text are required.

Recently a new proposal for the detection and guidance of online students who become disengaged has been introduced. This proposal is based on the principles of the Open Learner Model [10], [11], [12], aiming to help students focus reflection on their learning and progress. Learner models are models of learners' knowledge, difficulties and misconceptions and are essential for an adaptive learning environment to behave different for different students. Learner models are usually accessible to the students they model. Open learner models are learner models that are accessible to the student being modelled and sometimes also to other users (e.g. peers, teachers, instructors, tutors). It has been argued that the act of viewing representations of students' understanding can raise their awareness of their developing knowledge and difficulties at the learning process [13].

Arroyo et al., in [5] argue that non-invasive interactions can change a student's engagement state. More specifically, they propose the use of an open learner model as a mean to guide students into reengagement. Through the open learner model, performance and progress charts accompanied by tips and encouragement are presented to students, aiming to reduce gaming, enhance learning, while at the same time generate a more positive perception to the system and of the learning experience.

In the same line of non-invasive interactions based on the principles of the open learner model, we propose the use of the open learner model, as a mean for the detection and guidance of online students who become disengaged. More specifically we extend the work of Arroyo et al. in [5] by including in the open learner model not only performance and progress charts, but also a representation of students' working behaviour.

The open learner model described in this work (OLM_SCALE) was developed in the frame of a web-based, adaptive, activity-oriented learning environment referred to as SCALE (Supporting Collaboration and Adaptation in a Learning Environment) [14].

In order to investigate the impact of OLM_SCALE in guiding/stimulating disengaged students to work in a more effective and engaged way we conducted an empirical study. The main research questions of the empirical study were: (i) can OLM_SCALE stimulate students to work in an effective and engaged way? and (ii) what is students' opinion about the effectiveness of OLM_SCALE in supporting the learning process in the context of an introductory to Informatics and Telecommunications course?

2 Open learner Model maintained in SCALE (OLM_SCALE)

OLM_SCALE combines and expands ideas coming from the areas of computer based interaction and collaboration analysis [15], [16] and open learner modelling. In particular, we collect raw data from students' interaction with the system using a set of indicators and visualize this information alongside with comparative information coming from selected peers aiming to support the learning process at awareness and metacognitive level. At awareness level, the value of the indicator is presented to the

student and at metacognitive level, the calibrated (through a predefined form) value of the indicator is presented to the student.
Specifically, we designed a set of indicators that focus on individual activities and reflect the structure of SCALE's educational material. The indicators aim to:

(i) *reflect student's knowledge by using skill meters (*metacognitive level): indicators for student's performance level at activity, sub-activity and question items level,

(ii) *offer comparison to peers views of the learner model data*: indicators for answers given by other peers (awareness level), the minimum, maximum and average performance level, calculated from all the students enrolled in the specific subject matter (metacognitive level), and

(iii) *present students' working behaviour* (awareness & metacognitive level): indicators for student's interactions with the system (received feedback components, activities/subactivities elaboration attempts, minimum, maximum and average knowledge level as well as average elaboration attempts).

OLM_SCALE follows the simple model representation, that uses skill meters, as the structure of SCALE's educational material is already hierarchical and more complex learner views would require the definition of additional relations between the various concepts. As stated in the work of [17] a simple representation of the open learner model data that uses skill meters can have positive effect on students' learning and metacognition. Moreover the simple skill meter representation of the open learner model data has found to be an adequate representation for sharing learner models with peers and instructors [18], [19].

3 Indicative screenshots of the OLM_SCALE

Figure 1 illustrates the main screen of OLM_SCALE for the concept of "Algorithms" in the context of the subject matter "Introduction to Informatics and Telecommunications". Four activities are included in the specific concept, i.e. *Definition of an Algorithm*, *Pseudocode*, *Sequential Search* and *Binary Search*. For each activity the following indicators are illustrated: (i) the current knowledge level (*metacognitive* - based on skill meters), (ii) the elaboration attempts (*awareness* - i.e. how many times the activity's questions have been submitted), (iii) the activity's status (i.e. how many subactivities have already been elaborated) (*awareness),* and (iv) the minimum, maximum and average knowledge level (*metacognitive*) and (v) average elaboration attempts (*awareness)*. Also, the learner model includes functionalities that allow students: (i) to choose whether the information held in the model will be visible from their co-students, and (ii) to select their preferred feedback types.
As can be seen in Figure 1, the specific student inspects his own model and compares his developing of understanding of the target concept to that of the two peers he has chosen to inspect their models. For each activity the following information is externalized: (i) the student's knowledge level, (ii) the minimum, average and maximum knowledge level, (iii) the activity's elaboration attempts (e.g. for the activity *Definition of an Algorithm* the specific student has attempted twice to elaborate the corresponding subactivities), (iv) the activity's status presented either by the fraction indicating elaborated subactivities / available subactivites (e.g. for the activity *Pseudocode* the specific student has elaborated correctly two out of the three available subactivities) or by a specific icon indicating that all the available subactivites have already been elaborated (e.g. for the activity *Definition of an Algorithm* the specific student has elaborated correctly all the available subactivities), (v) the activity's average elaboration attempts and (vi) the peers' knowledge level and the corresponding elaboration attempts and status. By pressing the *Open/Close* button, the student can choose either to open or to close the model to his/hers peers. Through the *Users* button the student can choose the peers whose models s/he likes to inspect.

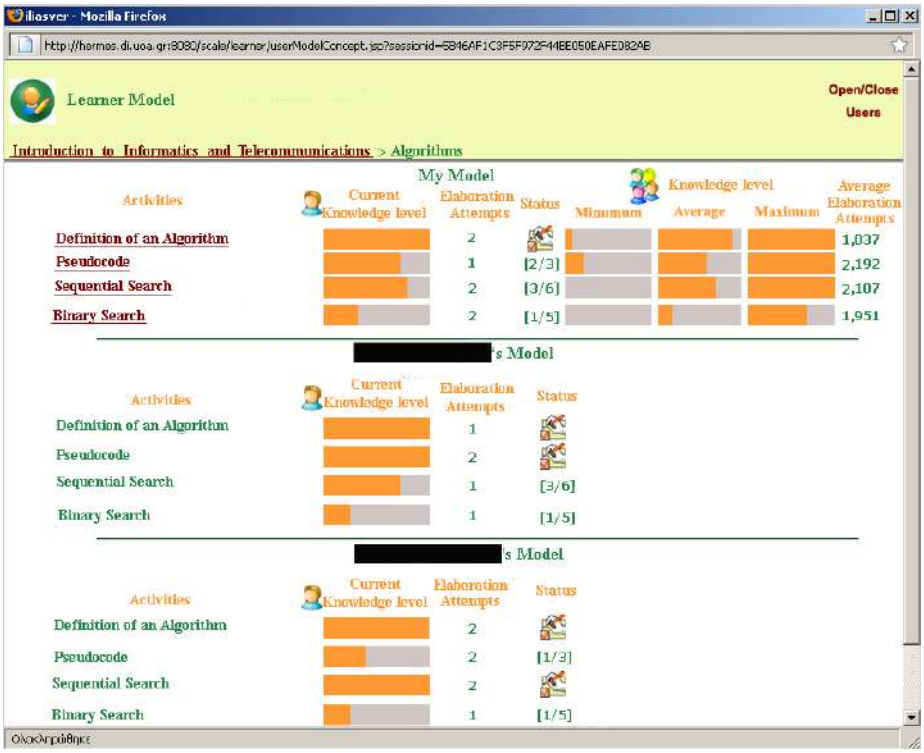


Figure 1: OLM_SCALE Screenshot of a specific learner-model's main screen of the concept "Algorithms" (translated in English)

4. The empirical study

During the academic year of 2007, SCALE and the corresponding educational material developed have been used for the first time aiming to improve the teaching and learning processes of the undergraduate course "Introduction to Informatics and Telecommunications" [20]. The course is compulsory and is taught 3 hours per week. The course objectives are as follows: (i) to give students a strong background in the following topics of computer science: Data Storage, Data Manipulation, Operating Systems, Networking and Internet, Algorithms and Programming Languages, (ii) to make students comfortable with computers and eliminate any fears about computers and (iii) to establish basic foundations of further study. The evaluation of SCALE's application during the academic year 2007 showed that although SCALE has been proved as a valuable tool in supporting and enhancing the teaching and learning processes, a considerable percentage of the participated students (27.8%) seemed rather disengaged while working in the environment. This fact encouraged us to develop an open learner model for SCALE and to conduct an empirical study in order to investigate the issue of guiding/tutoring the disengaged students to work in a more effective and engaged way. The empirical study was conducted during the winter semester of the academic year 2008-2009 in the context of the aforementioned course "Introduction to Informatics and Telecommunications" [21].

The main research questions of the empirical study was: Can the open learner model embedded in the SCALE environment stimulate students to work in an effective and engaged way?

154 first year's students, that enrolled to the course "Introduction to Informatics and Telecommunications" at the Department of Informatics and Telecommunications of the National and Kapodistrian University of Athens, participated in the study. All participants aged between 18 and 23 yeas attended General Lyceum in Greece during their Secondary Education years.

In order to investigate the effectiveness of the open learner model in the context of the specific course, educational material in the form of individual activities was developed. This material exploits the learning design of the SCALE environment and can be used (i) by the teacher as laboratory based exercises or as homework, and (ii)

by the student as a mean to deepen his/hers knowledge in the underlying topics or prepare him/herself for the corresponding university courses [22], [23], [24].
During the first lecture of the course the two responsible teachers presented an outline of the covered topics. Following that, one of the teachers presented the SCALE environment, the developed educational material and the open learner model. Results (see section Results) were obtained from system logs and questionnaires.
The eight weeks empirical study consisted of the following phases:

(i) *Pre-test* (1st week) – lasted 1,5 hour: All students participating in the empirical study, took the pre-achievement test.

(ii) *Working out activities* – lasted 7 weeks (1st week – 7th week): The participating students worked out the activities embedded in the SCALE environment. It was suggested to (a) access the environment and work out the corresponding activities at a week's basis following the material of the lectures and (b) use the open learner model. The estimated weekly time that students had to work with SCALE was 2 hours. At the end of the 3rd week a log file analysis was performed, in order to reveal (a) the students that were engaged / disengaged in working out the activities and (b) the students that used/did not use the open learner model. The students who did not use the open learner model were prompted (via email) to do so. The same log file analysis was repeated at the end of the 7th week of the empirical study.

(iii) *Post-test* (8th week) – lasted 3 hours: All students took the post-achievement test (course's final exam).

(iv) *Filling the questionnaire* (8th week) – lasted 30 minutes: The participating students were asked to express their opinion concerning the open learner model.

All students attended the weekly lectures and studied the relevant educational material (course book and lecture notes), in order to prepare themselves for the final exams. The lecture notes and supplementary material (e.g. announcements concerning the course, answers to questions posted during the lecture) were delivered to students through the course management system (http://eclass.di.uoa.gr/).
Moreover, educational material in form of individual activities was developed and delivered through the SCALE environment, covering the following topics: (i) Data Storage, (ii) Data Manipulation, (iii) Operating Systems, (iv) Networking and Internet, and (v) Algorithms. Each activity consisted of one or more sub-activities; and each sub-activity of one or more question items. The activity/sub-activity addressed learning outcomes of the Comprehension and/or the Application level. The question items were (i) multiple choice questions with one correct answer, (ii) multiple choice questions with more than one correct answers, (iii) matching questions, (iv) two-tier questions, where the second tier explores students' reasons for the choice made in the first tier [25], [26], and (iv) open answer questions (assessed by the teacher). The activities under consideration cover all difficulty levels, provide multiple and different kind of feedback types and are automatically assessed by the system (except of the open answer questions).
During the first week of the course all students participated in the pre-test, in order to identify their prior knowledge (10 multiple choice and 5 open answer questions). Each question scored 10 points.
During the last week of the course, all students participated in the course final exam (post-test). The post-test aims to reveal the differences in students' conceptions with the pre-test, after their involvement with the weekly lectures, the course educational material and the SCALE environment. The students had to answer the same questions they worked out in the pre-test. The evaluation of both the pre-test and the post-test was performed by the two course teachers in a 10-point scale (1-10) for each question. The final score of each question was the mean of the two evaluators' scores.

5. Results

To determine whether the open learner model maintained in SCALE environment stimulate students to work in an effective and engaged way, we performed (during the 3rd week of the empirical) a log file analysis in order to reveal the disengaged students. This analysis was based on the actions that students mostly performed

before reaching the correct answers of an activity, after initially submitting wrong answers.

In SCALE environment, whenever a student submits a wrong answer s/he has the possibility to identify and correct his/her errors (e.g. by receiving tutoring feedback components or by restudying the corresponding topic of the course book) and then resubmitting the answer.

The log file analysis revealed that some students had extremely low *resubmitting time* (i.e. the time elapsed between the initially wrong submission and the subsequent resubmission) and considerable high average rate of activities' elaboration attempts. Accordingly to the works of [6] and [27], we presume that these students were rather disengaged while working in the environment and were only trying to guess the correct answer, in contrast with the rest of the students, that before resubmitting the answer either restudied the question or consulted relevant educational material or received relevant tutoring feedback components.

In order to identify the presumed disengaged students, we calculated (i) the elapsed time between the initial submission of a wrong answer and the final correct submission the same question and (ii) the estimated time for random submission of answer to the specific question. For example, if the estimated time for random submission for a specific activity was 10 sec and a student resubmitted in less than 10 sec then this attempt to answer the question is considered a blind guess and the student rather disengaged when answering the specific question. The estimated time for random resubmission was calculated from the data derived by the two course teachers that deliberately tried to submit blind guesses to each activity's questions as quickly as possible until they reached the correct answer. Comparing these times (the actual resubmission time and the estimated time for random resubmission), we divided students in two subgroups:

- *Disengaged:* students that retry to answer the question very rapidly (less than the estimated time).
- *Engaged:* students that retry to answer the question after a considerable time interval (equal or more than the estimated time) or after receiving tutoring feedback components.

The next step was to classify each student according to their resubmitting time as *Engaged* or *Disengaged* by using Two-Step cluster analysis.

Moreover, we classified each student according to the extend s/he used the open learner model. More specifically, we calculated for each student the rate $Mu=T_M/T_T$ (*Model Usage*), where T_T represents the Total Time spent working in SCALE environment up to the end of the 3rd week of the empirical study and T_M represents the time spent using the open learner model during the same period. Out of all the occurred Mu vales, the value Mu_{30} was calculated, that corresponds to the specific Mu value of which 30% of the students have smaller Mu value. This way the students were divided in two subgroups:

- *Non Model Users* ***(NMU)****: students with* $Mu <= Mu_{30}$
- *Model Users* ***(MU)****: students with* $Mu > Mu_{30}$

The results of the students' classification according to their way of working (engaged – disengaged) and according to the extent they used the open learner model are presented in Table 1. As can be seen in Table 1 a great percentage of the **disengaged** students were Non Model users (80.8%) and a great percentage of the **engaged** students were Model Users (81.5%). This fact was an indication that the exploitation of the facilities offered by the open learner model can stimulate students to work in an engaged way. In order to verify this assumption, we encouraged the students who did not use the open learner model to do so, by informing them (via email) that using the open learner model might enhance and support their learning. More specifically, these students were encouraged to access the learner model, to make the information held in it available to their peers (i.e. *open* their model) and to choose their peers whose models they would like to inspect.

Table 1: Classification of students according to their way of working (Engaged – Disengaged) and to the extent they used the open learner model (Non Model Users – Model Users) at the end of the 3rd week of the empirical study

Model Usage / Way of working	**M**odel **U**sers	**Non M**odel **U**sers
Disengaged	14 (19.2%) Type *D_MU*	59 (80.8%) Type *D_NMU*
Engaged	66 (81.5%) Type *E_MU*	15 (18.5%) Type *E_NMU*

At the end of the 7th week of the empirical study we repeated the aforementioned log file analysis in order to investigate whether the 3rd week's email intervention would result on the students' more engaged way of working. The analysis revealed that several changes of students' types had occurred. More specifically we registered the students' change of types. The results are shown in Table 2:

Table 2: Registered changes of students' types (3rd vs. 7th week of the empirical study) – ranked according to occurrences

TYPE 3rd week →7th week	**DESCRIPTION** This category includes students that at the end of the… 3rd week of the empirical study were characterized as:	7th week of the empirical study were characterized as:	**STUDENTS**	**GROUP**
E_MU → *E_MU*	model users and worked in an engaged way	model users and worked in an engaged way	64	Group 1
D_NMU → *E_MU*	non_model users and worked in a disengaged way	model users and worked in an engaged way	37	Group 2
D_NMU → *D_MU*	non_model users and worked in a disengaged way	model users and worked in a disengaged way	13	Group 3
D_MU → *D_MU*	model users and worked in a disengaged way	model users and worked in a disengaged way	12	Group 4
E_NMU → *E_MU*	non_model users and worked in an engaged way	model users and worked in an engaged way	8	Group 5
D_NMU → *D_NMU*	non_model users and worked in a disengaged way	non_model users and worked in a disengaged way	7	Group 6
E_NMU → *E_NMU*	non_model users and worked in an engaged way	non_model users and worked in an engaged way	5	Group 7
E_MU → *D_MU*	model users and worked in an engaged way	model users and worked in a disengaged way	2	Group 8
E_NMU → *D_MU*	non_model users and worked in an engaged way	model users and worked in a disengaged way	2	Group 9
D_MU → *D_NMU*	model users and worked in a disengaged way	non_model users and worked in a disengaged way	2	Group 10
D_NMU → *E_NMU*	non_model users and worked in a disengaged way	non_model users and worked in an engaged way	2	Group 11

As can be seen in Table 2, 60 out of 74 students (81%) of the initially non model users (74 students – see Table 2) responded positively to our suggestion to use the open learner model (60 students – see Table 2 Groups 2, 3, 5, 9). More specifically,

37 out of 59 initially disengaged - non model users not only became Model users, but also managed to improve their way of working towards a more engaged way (Table 2 - Group 2). 13 out of 59 initially disengaged - non model users (22%) although responded positively to our suggestion to use the open learner model continued to work in an disengaged way (Table 2 – Group 3). 7 out of 59 initially disengaged - non model users (11.8%) avoided to use the open learner model continued working in a disengaged way (Table 2 – Group 6). Finally, 2 out of 59 initially disengaged - non model users (3.3%) avoided to use the open learner model but managed to improve their way of working towards a more engaged way (Table 2 – Group 11).
The majority (97.5%) of the initially model users (80 students – see Table 1) continued to use the open learner model till the end of the experimental study (78 students – see Table 2 Groups 1, 4, 8). The reader may notice that almost all initially type E_MU students (Engaged - Model users) remained in the same category throughout the duration of the empirical study.

Moreover, we examined the performance differences (pre-test vs. post-test), regarding the groups that contained more than 10 students (i.e. four groups: Group 1 (64 students), Group 2 (37 students), Group 3 (12 students), Group 4 (13 students)). No significant difference was found in the One-Way ANOVA between the four groups on the pre-test performance. (see Table 3 – pre-test column). As can be seen in Table 3 the four groups where initially equivalent in their pre-test performance ($p>0.05$).

Table 3: Evaluation of the pre-test and post-test – Between groups (Group 1, Group 2, Group 3, Group 4) One-Way ANOVA

pre-test		post-test	
F(3)	*p*	F(3)	*p*
0.647	0.899 (ns)	18.324	<0.01

As can be seen in Table 3 (post-test column), the results of the One-Way ANOVA revealed that the mean post-test performances are not equal across the groups and that at least one of the group means is significantly different from at least one other group mean ($p<0.05$). In other words the fact that the significance value of the F test is less than 0.05 suggests that the mean post-test performances of the four groups differ in some way. In order to obtain which groups are different and which are not, we performed LSD Multiple Comparison tests based on the four groups (see Table 4)

Table 4: One-Way ANOVA Multiple Comparison between groups (Group 1, Group 2, Group 3, Group 4)

(I) Type	(J) Type	Mean Difference (I-J)	sig
Group 1 (64 students)	*Group 2*	0.39	0.07 (ns)
	Group 3	2.02	<0.01
	Group 4	1.65	<0.01
Group 2 (37 students)	*Group 1*	-0.39	0.07 (ns)
	Group 3	1.62	<0.01
	Group 4	1.26	<0.01
Group 3 (13 students)	*Group 1*	-2.02	<0.01
	Group 2	-1.62	<0.01
	Group 4	-0.36	0.39 (ns)
Group 4 (12 students)	*Group 1*	-1.65	<0.01
	Group 2	-1.26	<0.01
	Group 3	0.36	0.39(ns)

As can be seen in Table 4, the mean of Group 1 is 0.39 points higher than the mean of Group 2. Since the significance level is larger (i.e. 0.07) than the required 0.05 alpha level, we conclude that the differences in post-test performance for Group 1 and the Group 2 are not significant. This fact suggest that the students of Group 2 although initially disengaged and non model users, (i) managed to improve their way of

working towards a more engaged way, (ii) became model users and (iii) improved their post-test performance at the same rate as the students that throughout the run of the empirical study were model users and worked in an engaged way (i.e. students of Group 1).
On the other hand the means of Group 3 and Group 4 are significantly lower from the means of the other groups. Specifically the students of Group 3 scored 2.02 & 1.62 points lower than the students of Group 1 and Group 2 respectively. The students of Group 4 scored 1.65 & 1.26 points lower than the students of Group 1 and Group 2 respectively. This fact suggests that the students of Group 3 and Group 4 remained disengaged while working in the environment and although they used the open learner model could not improve their post-test performance at the same rate as the students of Group 1 and Group 2. This could be explained due to the fact that these students tried to elaborate the requested activities either by trying to blind guess the correct answers or by trying to guess the correct answers through the open learner model.

The empirical study, that was conducted, showed that the exploitation of OLM_SCALE can guide the online students who become disengaged towards reengagement. More specifically although 38% of the participated students initially did not use OLM_SCALE and worked in a rather disengaged way, 63% out of these students not only managed to improve their way of working towards a more engaged way after they were prompted to use OLM_SCALE but also improved their post-test performance at the same rate as the students that throughout the run of the empirical study were model users and worked in an engaged way. This fact suggests that non invasive interactions based on the principles of the open learner model can help students focus reflection on their learning process and coax them to reengagement. It seems that including in the open learner model performance and progress charts, and a reflection of the students' working behaviour can effectively lead disengaged students to work in an engaged way. In the frame of this empirical study no further tips or encouragement were given to the participating students. Our results are in line with the results of [5] in terms that disengaged students became engaged by accessing the open learner model. But, our works goes one step further showing that additional tips (that may require skilfully log file analysis) and encouragement are not necessary (as reported in the work of [5]); students may think of their interaction and reflect both on their own working behaviour as well as on their co-students' and friends' working behaviour. The log file analysis performed in the frame of the empirical study presented in section 4 has been conducted in order to investigate whether OLM_SCALE can stimulate students to work in an effective and engaged way. Our results show that simply through the exploitation of the facilities offered by OLM_SCALE and without performing any log file analysis, disengaged students can be guided into reengagement. 22% out of the students that initially did not use OLM_SCALE and worked in a rather disengaged way, although responded positively to our suggestion and used OLM_SCALE continued to work in a rather disengaged way. These students improved their post-test performance at a significantly lower rate as the students that either throughout the run of the empirical study were model users or became model users after our suggestion. Moreover, 12% out of the students that initially did not use OLM_SCALE and worked in a rather disengaged way chose not to use OLM_SCALE even after our suggestion, and continued working rather disengaged till the end of the empirical study. Finally 3% out of the students that initially did not use OLM_SCALE and worked in a rather disengaged way although did not respond positively to our suggestion and chose not to use OLM_SCALE, manage to work in an engaged way at the end of the empirical study.
A considerable percentage of the students (43%) chose by themselves to use OLM_SCALE from the beginning of the study and were found to work in a rather engaged way. 97% out of these students continued to use OLM_SCALE and to work in a rather engaged way till the end of the empirical study. The rest 3% out of these students although continued to use OLM_SCALE were found to work rather disengaged at the end of the empirical study.
The participated students expressed their satisfaction regarding SCALE environment and in particularly they characterised the open learner model maintained in SCALE as a valuable and supportive mean in learning.

References

1. Montelpare, W.J., & Williams, A. (2000). "Web-based learning: challenges in using the Internet in the undergraduate curriculum", Education and Information Technologies, Vol. 5 No.2, pp.85-101.

2. Brusilovsky, P., & Peylo, C. (2003) 'Adaptive navigation and intelligent web-based educational systems', *International Journal of Artificial Intelligence in Education*, Vol. 12, Nos. 2-4, pp. 159-172.

3. Reimann, P., Freebody, P., Hornibrook, M., & Howard, S. (2009). Immersive learning environments: A study of teachers' innovation using The Le@rning Federation's digial learning resources., (retrieved from

   http://www.thelearningfederation.edu.au/verve/_resources/Study_of_teachers_using_TLF_resources.pdf)

4. OECD. (2009). Creating effective teaching and learning environments. First results from TALIS: OECD. (Retrieved from http://www.oecd.org/dataoecd/17/51/43023606.pdf)

5. Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B.P. (2007). Repairing Disengagement with Non-Invasive Interventions. *Proceedings of the 13th International Conference of Artificial Intelligence in Education*. IOS Press.

6. Cocea, M., & Weibelzahl, S. (2007). Eliciting motivation knowledge from log files towards motivation diagnosis for Adaptive Systems. In C. Conati, K. McCoy & G. Paliouras (Eds.), User Modeling 2007. *Proceedings of 11th International Conference, UM2007, Lecture Notes in Artificial Intelligence LNAI 4511 (© Springer Verlag)* (pp. 197-206). Berlin: Springer

7. Baker, R. S. J., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., & Beck, J. E. (2006). Adapting to when students game an intelligent tutoring system. Lecture Notes in Computer Science, 4053, 392.

8. Johns, J., & Woolf, B. (2006). A Dynamic Mixture Model to Detect Student Motivation and Proficiency. In: Proc of AAAI06, pp. 163-168. AAAI Press, Menlo Park

9. Baker, R. S. J., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., & Beck, J. E. (2006). Adapting to when students game an intelligent tutoring system. Lecture Notes in Computer Science, 4053, 392.

10. Bull, S., & Kay, J. (2005). A Framework for Designing and Analysing Open Learner Modelling, Proceedings of Workshop on Learner Modelling for Reflection, International Conference on Artificial Intelligence in Education 2005, 81-90.

11. Bull, S., & Kay, J. (2007). Student Models that Invite the Learner In: The SMILI Open Learner Modelling Framework, International Journal of Artificial Intelligence in Education 17(2), 89-120.

12. Bull, S., & Britland, M. (2007). 'Group Interaction Prompted by a Simple Assessed Open Learner Model that can be Optionally Released to Peers', Workshop on Personalisation in learning environments at individual and group level, UM07, Corfu.

13. Bull, S. Brna, P., & Pain, H. (1995), Extending the Scope of Student Models, *User Modeling and User Adapted Interaction,* v. 5(1), pp 45-65.

14. Gogoulou, A., Gouli, E., Grigoriadou, M., Samarakou, M., & Chinou, D. (2007). A web-based educational setting supporting individualized learning, collaborative learning and assessment. *Educational Technology & Society Journal*, 10(4), 242-256.

15. Dimitracopoulou, A., Martinez, A., Dimitriadis, Y., Morch, A., Ludvigsen, S., Harre, A., Hoppe, U., Barros, B., Verdejo, F., Hulsof, C., de Jong, T., Fessakis, G., Petrou, A., Lund, K., Baker, M., Jermann, P., Dillenbourg, P., Kollias, ZV, & Vosniadou, S. (2005). State of the Art of Interaction Analysis for Metacognitive Support & Diagnosis. Deliverable 31.1.1, Interaction Analysis JEIRP, Kaleidoscope Network of Excellence.

16. Reimann, P. (2003). How to Support Groups In Learning: More Than Problem Solving. Invited talk. In Supplementary Proceedings of the 11th International Conference on Artificial Intelligence in Education, AIED 2003, Sydney, 3-16.

17. Mitrovic, A., & Martin, B. (2007). Evaluating the Effect of Open Student Models on Self-Assessment. International Journal of Artificial Intelligence in Education, 17(2).

18. Bull, S., Mabbott A., & Abu Issa, A.S. (2007). UMPTEEN: Named and Anonymous Learner Model Access for Instructors and Peers. International Journal of Artificial Intelligence in Education, 17(3).

19. Lazarinis, F., & Retalis, S. (2007). Analyze Me: Open Learner Model in an Adaptive Web Testing

System. International Journal of Artificial Intelligence in Education, 17(3).

20. Verginis I., Gogoulou A., Gouli, E., Boubouka M., and Grigoriadou M. (2009). Enhancing Learning in Introductory Computer Science Courses through SCALE: An empirical study. *IEEE Transactions on Education* , 54, 1, 1-13.

21. Verginis I, Gouli E., Gogoulou A., and Grigoriadou M. (2010). Guiding learners into Reengagement through SCALE environment: An empirical study, *IEEE Transactions on Learning Technologies 2010* (under publication).

22. Verginis I., Gogoulou A., Gouli E., Grigoriadou M. (2008). Supporting Learning in Introductory Computer Science Courses through the SCALE Environment. *Proccedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA 2008)*, Vienna, Austria, 3313-3318

23. Γρηγοριάδου, Μ., Βεργίνης Η., Γόγουλου Α., και Γουλή Ε. (2009) Υποστήριξη της διδακτικο-μαθησιακής διαδικασίας με δραστηριότητες μέσω του μαθησιακού περιβάλλοντος SCALE. Στο Μ. Γρηγοριάδου, Ε. Γουλή, Α. Γόγουλου (επιμ.) *Διδακτικές Προσεγγίσεις και Εργαλεία για τη διδασκαλία της Πληροφορικής*, Εκδόσεις Νέων Τεχνολογιών, 519 – 554.

24. Βεργίνης Η., Μπούμπουκα Μ., Γόγουλου, Α. Γουλή, Ε. (2009). Καθοδηγώντας τους μαθητές κατά την εκπόνηση δραστηριοτήτων με πολλαπλές μονάδες ανατροφοδότησης μέσω του περιβάλλοντος SCALE . *Πρακτικά 5ου Πανελλήνιου Συνεδρίου των Εκπαιδευτικών για τις ΤΠΕ*

25. Tsai, C.-C., & Chou, C. (2002). Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted Learning*, 18, 157-165.

26. Tsaganou G., & Grigoriadou M., "Authoring with ReTuDiSAuth for Adaptive Learning from Text". (2009). The International Journal of Learning, 16 (10), 1-10, http://www.Learning-Journal.com, ISSN 1447-9494

27. Beck, J. (2005). Engagement tracing: using response times to model student disengagement. *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, p. 88–95.

A Web-based Environment for Supporting Learning Communities in Distance Education – Use in Computer Science Education

Stefanos Ziovas[1]

National and Kapodistrian University of Athens

Department of Informatics and Telecommunications

sziovas@di.uoa.gr

Abstract. In the context of this dissertation, issues related to design and development of educational environments that support collaborative learning and contribute to the active engagement of learners in knowledge building are studied. Specifically, the dissertation is placed in the context of supporting the learning process by utilizing the Internet and especially the Environments which support Electronic Learning Communities by leveraging social software technologies for access to information and knowledge creation. In the frame of this thesis the development of innovative techniques to support the creation, management, operation and interconnection of electronic communities was studied. A Web based educational environment CRICOS (CReate Interconnected COmmunitieS) was designed and implemented, based on the theory of learning communities and particularly communities of practice, which supports members of different communities in collaborative knowledge creation

Keywords: Learning Communities, Communities of Practice, Social Tagging, Collaborative Filtering, Social Navigation

1. Introduction

Online communities provide the ground for knowledge creation, social negotiation of meaning and learning through the collective participation of their members. In order to participate effectively into these processes community members need guidance to find and synthesise information. Over the past few years several community-driven

[1] Dissertation advisor: Maria Grigoriadou, Professor

technologies have been developed to address this issue [1]. The role of these technologies is dual fold in the sense that they focus on creating communities by bringing people together and also they exploit the community around an individual as a valuable resource to guide information seeking.

Collaborative filtering systems recommend data items to a user by taking into account the opinions of other users [2]. Instead of recommending data items because they are similar to items the user preferred in the past (content-based recommendations) collaborative approaches generate recommendations about data items that users with similar interests liked in the past [3]. Social navigation systems help users to navigate in a community information space by making the aggregated behaviour of the community visible. The concept was introduced by Dourish and Chalmers as moving "towards" a cluster of other people or selecting objects because others have been examining them [4]. Social tagging is one of the latest popular approaches for information management and sharing. The approach became popular on the Web as part of the social bookmarking systems [5]. Social bookmarking systems allow users to create personal collections of bookmarks and share their bookmarks with others. Moreover users of these systems may organize their collection by entering keywords which are meaningful to them. This type of manual indexing is called tagging with index terms referred to as tags.

Several of the above mentioned technologies have been applied in the field of e-learning. CoFIND [6], [7], is a resource sharing system which allows students to provide feedback about resources and classifying them using several topics. The system encourages not only the tagging of resources according to topics, but also with pedagogical metadata known as qualities, which are then used to supply ratings on the resources. EDUCO [8] is a collaborative learning environment which visualizes the information space as clusters of documents. The documents change their color according to how much they have been read in relation to other documents. Live users presented in the learning environment as coloured dots located next to the documents that they are currently viewing. The system supports both synchronous and asynchronous communication among users. Knowledge Sea II [9] focuses on helping students of introductory programming courses to find relevant readings among hundreds of online tutorial pages distributed over the Web. The system represents clusters of pages in varying shades of blue by taking into account both number of visits and time spent on every page.

In the context of facilitating knowledge creation among the members of different communities with the use of community-driven technologies the design of a web-based environment is proposed, referred to as CRICOS (CReating Interconnected COmmunitieS) [10], [11], [12], which supports knowledge sharing within and across communities. The information space of each community is structured as a semantic network of nodes such as thematic categories, concepts, resources or members of the community. The users instead of tagging resources with keywords they can relate them with one of the existing nodes. Moreover, the web-based environment incorporates communication facilities to enhance the dialogue among the members of a community. The CRICOS environment incorporates a recommendation mechanism which is based on the integration of collaborative filtering and social navigation approaches to support community members to find the appropriate information. The first results from a study are encouraging regarding the usability and usefulness of the

provided facilities and revealed the users positive attitude to the CRICOS environment.

2. The CRICOS Environment

The CRICOS web-based environment was constructed with the aim of improving knowledge creation and sharing among the members of communities. Three main features of the environment support this:

- A flexible information space which is designed so as to maximize the ability of the members of a community to create and improve both its content and its organization.
- Facilities that support the discourse in the community by enhancing the communication among its members
- A recommendation mechanism based on collaborative filtering and social navigation approaches, which helps the members to find the appropriate information.

2.1 The Information Space of CRICOS

With the aim to move towards more community organized information spaces, in CRICOS the information space of each community is structured as a semantic network. Nodes represent thematic categories, concepts, resources or members of a community. Edges represent the relationships of nodes and can be labeled with a type. Members of a community structure the information space by contributing and placing new information in relation with information that is already presented in the information space. Thus, the information space evolves with the evolution of the community. This design approach structures the information space as a network of information nodes and creates very rich navigation opportunities [13].

The domain of a community in CRICOS is represented as a hierarchy of thematic categories and a semantic network of concepts. The first forms the common taxonomy of the community and is maintained by experienced members called "category editors". The role of a Category editor is given to members who have a certain level of knowledge of the thematic category (in the case of a course this role could be given to the teachers or tutors) and thus are responsible for the relevance of resources and concepts that are classified in their thematic category. The semantic network of concepts is constructed collaboratively by the members of the community who introduce concepts and their relationships. Instead of tagging resources a member can relate his resource with the available concepts. In the case that the resource cannot be related with the available concepts the member can introduce a new one. This approach provides to the community a certain level of flexibility upon the classification of the shared resources. A community can be based for the classification of resources on a defined taxonomy of thematic categories, on a folksonomy of concepts introduced by the members or on a mixed approach which synthesize taxonomy and folksonomy.

Resources may be located locally or externally in the WWW. A member who contributes a resource categorizes it according to one of the thematic categories. Moreover, the contributor may semantically link the submitted resource with other resources, concepts, thematic categories or members of the community. In order to

provide a semantic link a member may choose among several types of semantic relationships that other members have created or create a new type according to his perspective. Thus a member can represent his perspective of the knowledge stored in the repository not only by providing the content of his resource but also by locating his resource in the information space. Consequently, the information space of the community becomes a combination of the multiple perspectives that members have upon the community's domain. This approach supports the construction of meaningful structure overlays on stored resources and provides to the end-user valuable contextual information.

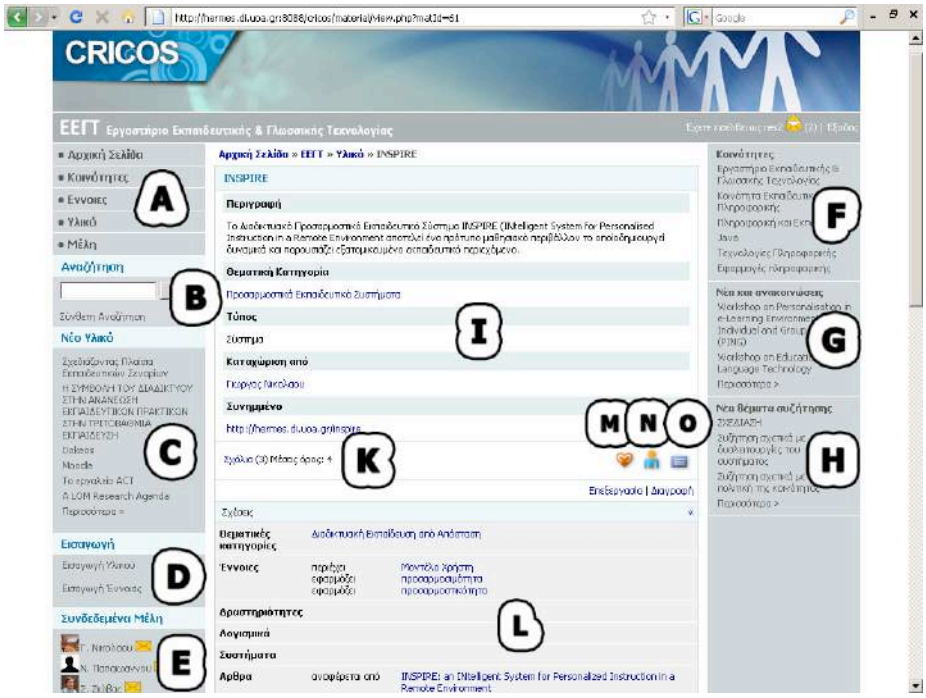


Fig. 1: A Snapshot from a Page Presenting Information about a Resource

In Figure 1 a screenshot from a page which displays information about a resource is presented. The left and right sidebars are always available to the end-user while navigating the information space. The left sidebar is divided into five blocks. The first block (A) provides links to the alternative starting points for navigation such as communities, concepts, resources and members. The second block (B) allows the user to search the information space. The third one (C) provides links to the recently uploaded resources. From the fourth block (D) the user can add a resource or a concept by following the relevant link. The last block (E) presents a list with the users who are online. The right sidebar contains one block (F) with links to the hosted communities, one (G) with the news and announcements and one (H) with the recently added discussion topics.

The main section provides useful information about the content and context of the resource. The upper side of the section (I) includes the title, the description and the type of the resource, the thematic category it belongs to, the name of the member who contributed the resource and a link to the resource itself. A link to the user comments and the average rating of the resource (K) is also available to the visitor. In order to provide information about the context of the resource the environment aggregates all the user-defined relations which include the resource and presents the links to the other relevant resources along with the type of the relation grouped by concepts, resources and members (L). For each link the type of the relation is also presented. For example a learning environment can be associated with several academic articles. The type "referred by" can denote articles which have references to the learning environment and the type "described in" articles which provide details about the learning environment. A user who discovers the resource can give his opinion on it by adding a comment and/or providing a numerical rating on a 1 to 5 rating scale (M). The user can also directly recommend the resource to one or more members who appear in his friends list (N). In this case each of the user's friends will receive a message with a link to the recommended resource. Moreover the user may add the resource to his personal collection and classify it in one of his familiar categories (O).

2.2 Communication Facilities

The environment supports the communication between its members via discussion forums and messages. Each discussion forum contains several discussion topics, organized into several threads. Each thread is hierarchically organized in a messages tree with the aim of presenting the overall structure of the conversation to the reader. A discussion forum can be attached to a community, a category or a resource. Members who are responsible for the above components have a moderator role in the attached discussion forums. With this approach even a simple member of the community may moderate a forum and engage in discussion with other members regarding the uploaded resource. In order to connect the parts of the dialogue with the community's domain we have introduced a mechanism which allows members to relate their post with the existing concepts. Thus when a member visits a concept he can have access not only to the related concepts and resources but also to the related posts. With this approach the browsing possibilities of the end users are broadened to both the content and the dialogue that took place in the community.

A messaging mechanism facilitates the communication among the users of the environment as well as the communication among the environment and the users. The message box of a user can receive messages from other users, friendship and group invitations, friends' recommendations and system messages. System messages include personalized system recommendations to the user e.g. a recommendation for a recently uploaded resource which matches user interests and requests for user approval on modifications that concern the user's model e.g. add a new thematic category to his interests.

2.3 Recommendation Mechanism

To allow users to obtain information that fits to their needs, knowledge or interests CRICOS keeps a user model for each user. The user model maintains and constantly

updates information that includes user membership and roles in communities and groups, user interests, user relationships and user activity. User interests are represented as a list of thematic categories in which the user is interested. User relationships are represented as a list of the users with whom the user has a relationship and for each user the type of relationship (e.g. friend, belonging to the same community, belonging to the same group). User activity is a log of the actions performed by the user.

Two main issues were taking under consideration while designing the recommendation mechanism. The first issue was concerning the calculation of the utility of a resource. Most of the collaborative filtering systems are based on explicit ratings (e.g. a 1-5 rating scale). This approach provides more accurate description of the user's opinion about a resource but requires extra effort from the user in order to provide the rating. Implicit ratings could be inferred by observing the user behaviour with a resource. However these approaches may lead to imprecise ratings.

The CRICOS environment follows a hybrid approach for recommending resources. The utility of a resource is calculated by aggregating both explicit and implicit ratings. Explicit ratings are provided by users who rate the quality of a resource on a 1 to 5 rating scale. Implicit ratings are inferred by observing the user behaviour with a resource and collecting the relevant data. Data from observations include the number of visits on the resource's page, if and how many times the user directly recommends the resource to others and if the resource is included in the user's personal collection. Both explicit and implicit ratings are combined into a single estimated rating which represents the utility of the resource for a particular user.

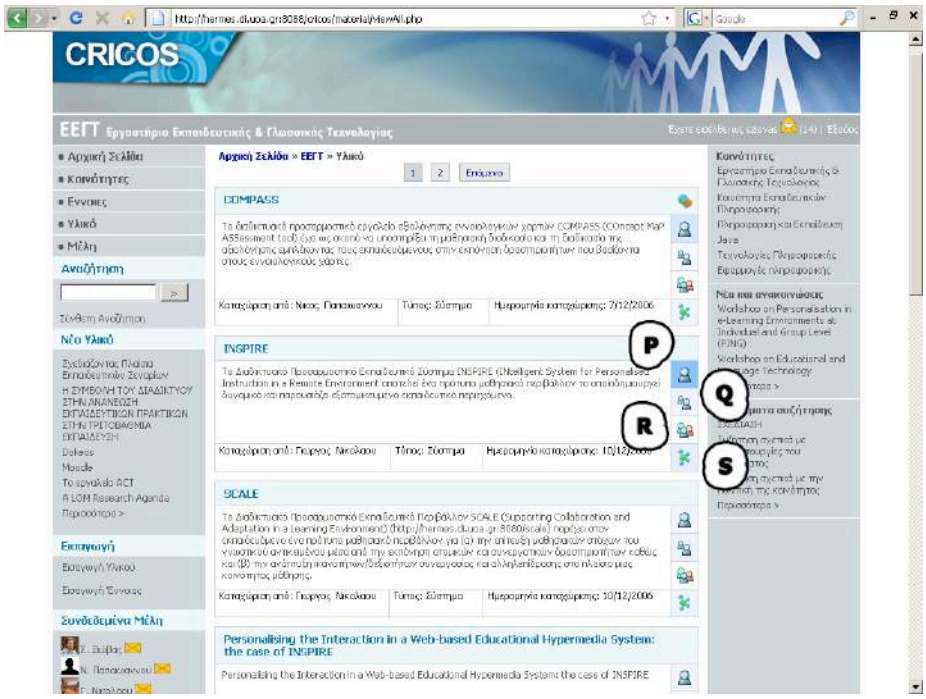


Fig. 2: A Snapshot from a Page for Browsing the Available Resources

Besides the utility of a resource for a single user, CRICOS averages the estimated ratings for two clusters of users that their opinions about the resource could be valuable for the user; user's friends and users with similar interests. The level of trust of the user on the system's recommendations was the second issue which was taking under consideration. Most of the systems that use collaborative filtering approaches they base their recommendations on the opinions of a group that has similar tastes with the user. The user should trust the recommendations without knowing the users of this group. In order to provide another hint on the utility of a resource besides the rating of the users with similar interests we provide also the rating of the user's friends.

CRICOS utilizes the above mentioned information to provide visual cues that help a user to find the most appropriate resources. For the visualization of the estimated ratings the environment follows an approach similar to that employed in Knowledge sea [9] For example in Figure 2 the environment presents the resources of a community and recommends relevant resources according to their utility for the user, for his friends and for other users who have the same interests. The estimated rating is represented by the background color of the box that includes the relevant icon: me (P), my friends (Q), users who have the same interests with me (R), the higher the rating the darker the color.

Regarding the quality of the available information, besides the provided rating mechanism we expect that the number of links to a resource will reflect its quality. This information is presented to the user when he browses the catalogue of resources

in the background color of the relationships icon (S), the higher the number of relationships the darker the color. On the same page an icon indicates whether a resource is a boundary object. Boundary objects are realised as resources or concepts of a community that have relationships with resources or concepts of another community. The relationships are provided from users who are members of these communities (brokers) and they are the contributors of at least one resource or concept that takes part in the relationship.

3. The Empirical Study

An empirical study was conducted aiming to investigate (i) the perceptions of students concerning the usability of CRICOS and (ii) the attitude of students towards community-driven technologies and the CRICOS environment. The empirical study took place in the context of the "Distance Learning" postgraduate course offered by the Department of Informatics and Telecommunications of the University of Athens. Fourteen postgraduate students participated in the study, which lasted eight weeks in total. The students where grouped into two groups. Each group used CRICOS in order to create a community on a domain of their choice and contribute educational material appropriate for distance learning. The project was carried out in three phases.

During the first phase (duration: 1 week) students had to submit a scenario concerning the framework of their work. The scenario included the domain of the community, the assigned roles of the members (e.g. category editors) and the types of the resources (e.g. activities, quizzes). One group decided to build a community around the domain of "Information Technologies" and the other around the domain of "Java Beans". First the scenario was uploaded to CRICOS, then it was discussed by the students on the attached discussion forum and based on this dialogue the scenario was revised. During the second phase (duration: 4 weeks) the students had to share their knowledge on the community's domain by exploiting the facilities of CRICOS. The students structured the information space of the community by adding the thematic categories, the concepts and the resources. During the third phase (duration: 3 weeks) the students of each group visited the community of the other group and acted as newcomers with the aim of learning the community's domain.

Upon the completion of the third phase, students were asked to fill and submit two questionnaires concerning the evaluation of the CRICOS environment. The first questionnaire was based on the Computer System Usability Questionnaire (CSUQ) [14]. The questionnaire consists of nineteen usability questions to which the respondent has to disagree or agree using a seven-point scale, ranging from 1 (strongly disagree) to 7 (strongly agree). The questionnaire has been extended with four additional elements in order to measure some specific characteristics of CRICOS (see Table 1, questions 9, 10, 18, 19). The second questionnaire was designed to obtain the participants' perceptions of the usefulness and usability of the specific features of CRICOS and it consisted of the following dimensions: structure and organization of the information space (11 items); social navigation facilities (12 items); member's information (5 items); communication facilities (12 items); recommendations provided by the system (8 items); facilities for interconnecting communities (7 items). The range of all the question items was from 1 to 5. Additionally open-ended questions were also included in the questionnaire to obtain users' comments and suggestions to improve the facilities of each dimension. A page

was reserved at the end of the questionnaire for students to provide general opinions and suggestions for the improvement of the system.

Discussion

The analysis of the responses of the CSUQ questionnaire showed that CRICOS rank high in terms of usability (Table 1). However there are some items that signal some improvements to the system. Question 11 (The system gives error messages that clearly tell me how to fix problems) has a mean score of 5 (std. dev. =1.47) which is the lowest score of the questionnaire. Relatively low is also the score of question 13 - The information (such as on-line help, on-screen messages and other documentation) provided with this system is clear – which has a mean score of 5.23 (std. dev. =1.09). These suggest that students need better guidance from the system and better suggestions to the errors they encountered during their work.

Table 1: Results Concerning the Usability of the CRICOS Environment (Scale 1-7)

No	**Question item**	Mean	**SD**
1	Overall, I am satisfied with how easy it is to use this system.	5,71	0,73
2	It is simple to use this system.	5,31	1,25
3	I can effectively complete my work using this system.	6,21	1,05
4	I am able to complete my work quickly using this system.	5,86	1,1
5	I am able to efficiently complete my work using this system.	5,17	1,27
6	I feel comfortable using this system.	5,93	1,38
7	It was easy to learn to use this system.	5,36	0,93
8	I believe I became productive quickly using this system.	5,21	1,63
9	**It is easy to communicate with other users of the system**	5,77	1,09
10	**The system facilitates the collaboration with other users**	6	1,24
11	The system gives error messages that clearly tell me how to fix problems.	5	1,47
12	Whenever I make a mistake using the system, I recover easily and quickly.	5,67	1,23
13	The information (such as on-line help, on-screen messages and other documentation) provided with this system is clear.	5,23	1,09
14	It is easy to find the information I need.	5,36	1,08
15	The information provided with the system is easy to understand.	6,14	0,95
16	The information is effective in helping me complete my work.	6,23	1,17
17	The organization of information on the system screens is clear.	5,64	1,39
18	**It was easy to receive information from other users of the system**	6,46	0,78

19	**The system allows me to share knowledge with other users**	6,69	0,63
20	The interface of this system is pleasant.	6,29	1,14
21	I like using the interface of this system.	6	1,3
22	This system has all the functions and capabilities I expect it to have.	5,54	1,2
23	Overall, I am satisfied with this system.	5,93	0,73

Some of the results from the analysis of the second questionnaire are presented in Table 2. From the results, it is obvious that the students found most of the provided facilities useful. A considerable number of students believed that a community's domain could be represented effectively in the CRICOS environment and the organization of the information space facilitates the access to the resources of a community and also the contribution of new resources. Regarding the recommendations provided by the environment, it seems that the students preferred more direct approaches for recommendations such as the possibility to receive systems messages whenever resources which are relative to their interests are uploaded (Question 9 – mean score = 4,79) than the indirect approaches of social navigation (Question 13 – mean score = 3,29, Question 14 – mean score = 3,5).

Table 2: Results Concerning the Attitudes of Students Towards the CRICOS Environment (scale 1-5)

No	**Question Item**	**Mean**	**SD**
1	Do you believe that a community's domain is represented effectively in the CRICOS environment? 1=not at all, 5=a lot	4,21	0,58
2	Do you think that the multiple perspectives which the members of a community have about the community's domain are effectively represented in the information space of CRICOS? 1=not at all, 5=a lot	4	0,78
3	Do you think that the organization of the information space in CRICOS facilitates the access to the information resources of a community? 1=not at all, 5=a lot	4	0,96
4	Do you believe that the organization of the information space facilitates the user to contribute information resources according to her knowledge and experience? 1=not at all, 5=a lot	3,69	0,75
5	The possibility to have discussion forum at a community is considered as … 1=not useful, 5=very useful	4,5	0,85
6	The possibility to have discussion forum at a thematic category is considered as …1=not useful, 5=very useful	4,29	0,91
7	The possibility to have discussion forum at a information resource is considered as …1=not useful, 5=very useful	4,64	0,5
8	The facility concerning the exchange of messages between users is considered as … 1=not useful, 5=very useful	4,57	0,85
9	The facility concerning the messages sent to a user by the system whenever new, relative to her interests, resources	4,79	0,43

	are uploaded is considered as … 1=not useful, 5=very useful		
10	The possibility to view the resources, concepts, thematic categories and members that other users consider as relevant with a resource you are viewing is considered as … 1=not useful, 5=very useful	4,57	0,65
11	The possibility to view the concepts that other users consider as relevant with the post you are reading is considered as … 1=not useful, 5=very useful	4,64	0,5
12	The information regarding your actions upon a resource as it is visualized with the relevant icon is considered as … 1=not useful, 5=very useful	3,57	1,02
13	The information regarding your friend's actions upon a resource as it is visualized with the relevant icon is considered as … 1=not useful, 5=very useful	3,29	1,07
14	The information regarding the actions upon a resource of the users who have similar interests with you, as it is visualized with the relevant icon is considered as … 1=not useful, 5=very useful	3,5	1,16
15	The information regarding the number of the relations of a resource as it is visualized with the relevant icon is considered as … 1=not useful, 5=very useful	4,07	0,92

As far as the analysis of the open questions is concerned, a considerable number of students expressed the opinion that the environment should incorporate facilities for synchronous communication. Also, 43% of the students expressed their need to be informed by email whenever new messages are received in their message box of the CRICOS web-based environment.

4. Conclusions and Future Plans

In the context of this thesis, we presented CRICOS, a web-based environment, which utilizes several community-driven technologies to support knowledge creation and sharing within interconnected communities. The environment incorporates a flexible information space which is designed so as to maximize the ability of the members of a community to create and improve both its content and its organization. Moreover, the environment provide communication facilities to support the discourse in the community and a recommendation mechanism, based on collaborative filtering and social navigation approaches, that helps community members to find the appropriate information. The empirical study we conducted in the context of the evaluation of the CRICOS environment reveals the positive attitude of the students towards the community-driven technologies and the CRICOS environment. The results concerning the usability of the environment and the facilities provided, indicating that CRICOS can support effectively the members of a community to share their knowledge about the community's domain. Our future plans include the improvement of the facilities provided according to the student's suggestions.

References

1. Ziovas S., Grigoriadou M. and Samarakou M. (2009). Supporting Learning in Online Communities with Social Software: an Overview of Community Driven Technologies, *Advanced Learning*, Raquel Hijon-Neira (Ed.), ISBN: 978-953-307-010-0, InTech

2. Schafer, J. B., Frankowski, D., Herlocker, J. & Sen, S. (2007). Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Neidl, W. (eds.) *The Adaptive Web: Methods and strategies of Web Personalization*. LNCS, vol. 4321, (pp. 291–324). Springer, Heidelberg

3. Burke, R (2007). Hybrid Web recommender systems. In: Brusilovsky, P., Kobsa, A., Neidl, W. (eds.) *The Adaptive Web: Methods and Strategies of Web Personalization*. LNCS, vol. 4321, (pp. 377–408). Springer, Heidelberg

4. Dourish, P., and Chalmers, M. (1994). Running out of Space: Models of Information Navigation, in *Proceedings of Human Computer Interaction*, Glasgow, Scotland.

5. Hammond, T., Timo, Hannay, Lund, B., and Scott, J. (2009). Social bookmarking tools (I), *D-Lib Magazine*, 11(4).

6. Dron, J., Boyne, C., Mitchell, R. & Siviter, P. (2000). CoFIND: steps towards a selforganising learning environment. In: Davies, G., Owen, C. (eds.) *Proceedings Of WebNet'2000, World Conference of the WWW and Internet* (pp. 75 – 80), San Antonio, Texas, US: AACE.

7. Dron, J. (2006). "The way of the termite: a theoretically grounded approach to the design of e-learning environments," *International Journal of Web Based Communities*, 2(1), 3 – 16.

8. Kurhila, J., M. Miettinen, P. Nokelainen, H. Tirri. (2002). EDUCO - A Collaborative Learning Environment Based on Social Navigation. In *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems* (pp. 242-252). Malaga, Spain: Springer, Verlag

9. Farzan, R. and Brusilovsky, P. (2005). Social navigation support through annotation-based group modeling. In: L. Ardissono, P.Brna and A. Mitrovic (eds.) *Proceedings of 10th International User Modeling Conference* (pp. 463-472). Berlin: Springer Verlag.

10. Ziovas S, Grigoriadou M. (2008). "Knowledge Sharing within and across communities: The CRICOS system", in Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies (ICALT2008), pp. 562-563, Cantabria, Spain

11. Ziovas S, Grigoriadou M. (2008) "CRICOS: A web-based system for creating interconnected communities", in *International Workshop on Social and Personal Computing for Web-Supported Learning Communities* (SPeL 2008) at the 2008 International Symposium on Applications and the Internet (SAINT 2008), pp. 313-316 Turku, Finland, Jul. 28- Aug. 1 2008

12. Grigoriadou, M. & Ziovas, S. (2006). Crossing the boundaries between communities of practice – A virtual environment for knowledge sharing. In E. Pearson & P. Bohman (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* 2006 (pp. 1568-1573), Chesapeake, VA: AACE.

13. Ziovas, S., Grigoriadou, M., & Samarakou, M. (2010) "Supporting Knowledge Sharing with Community-Driven Technologies: The Case of CRICOS", *The International Journal of Learning*, 2010, Vol. 17, Issue 1, pp 397-410.

14. Lewis, J.R. (1995): IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7(1), pp. 57–78.