15TH ANNUAL EUROMEDIA CONFERENCE

2009

BRUGES, BELGIUM

APRIL 15-17, 2009

Organized by

ETI

HASSELT UNIVERSITY

Sponsored by

EUROSIS

TU DELFT

TTVI

EU-DG INFSO

BELGACOM

GHENT UNIVERSITY

HOSTED BY

NOVOTEL BRUGGE CENTRUM

EUROMEDIA'2009

FEATURING

FIFTEENTH ANNUAL SCIENTIFIC CONFERENCE
ON WEB TECHNOLOGY, NEW MEDIA
COMMUNICATIONS AND TELEMATICS THEORY
METHODS, TOOLS AND APPLICATIONS

Jeanne Schreurs

APRIL 15-17, 2009
BRUGES, BELGIUM

A Publication of EUROSIS-ETI

Printed in Ghent, Belgium

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

Editor

Jeanne Schreurs
Hasselt University
Diepenbeek, Belgium

Programme Committee

WEBTEC Programme Committee

Dr. Paul Dowland, University of Plymouth, Plymouth, United Kingdom
Jens Lichtenberg, Ohio University, Athens, USA
Prof. Dr. Jörn Loviscach, Fachhochshule Bielefeld, Bielefeld, Germany
Assoc. Prof. Wenji Mao, Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China
Lorenzo Motta, Ansaldo Segnalamento Ferroviaro s.p.a. Genova, Italy
Dr. H. Joachim Nern, Aspasia Knowledge Systems, Dusseldorf, Germany
Dr. Carlos E. Palau, Universidad Politecnica de Valencia, Valencia, Spain
Prof. Paola Salomoni, Universita di Bologna, Bologna, Italy
Dr. Matthew Warren, Deakin University Geelong, Victoria, Australia

MEDIATEC Programme Committee

Assoc. Prof. Vincent Charvillat, IRIT-ENSEEIHT, Toulouse cedex, France
Dr. Fernando Boronat Segui, Universidad Politecnica de Valencia, Gran de Gandia, Spain
Jehan Francois Paris, University of Houston, Houston, USA
Prof. Marco Roccetti, Universita' di Bologna, Bologna, Italy
Dr. Leon Rothkrantz, Delft University of Technology, Delft, The Netherlands
Dr. Leonid Smalov, Coventry University, Coventry, United Kingdom

COMTEC Programme Committee

Prof. Dr. Marwan Al-Akaidi, De Montfort University, Leicester, United Kingdom
Boguslaw Butrylo, Bialystok Technical University, Bialystok, Poland
Dr. Nathan Clarke, University of Plymouth, Plymouth, United Kingdom
Dr. Steven Furnell, University of Plymouth, Plymouth, United Kingdom
Prof. Chris Guy, The University of Reading, Reading, United Kingdom
PhD Mohammad Riaz Moghal, Ali Ahmad Shah-University College of Engineering and Technology, Mirpur, Pakistan.
Ph. D. Oryal Tanir, Bell Canada, Montreal, Canada

INTERNATIONAL PROGRAMME COMMITTEE

APTEC Programme Committee

Prof. Dr. J. Broeckhove, RUCA-UA, Antwerp, Belgium
Dr. Juan Carlos Guerri Cebollada, Universidad Politecnica de Valencia, Valencia, Spain
Dr.ir. Johan Opsommer, Belgacom - BUS, Brussels, Belgium
Prof. Jeanne Schreurs. Hasselt University, Diepenbeek, Belgium
Ass. Prof. Ramiro Velázquez, Universidad Panamericana, Aguascalientes, Mexico
Dr. Charles van der Mast, Delft University of Technology, Delft, The Netherlands

E-TEC Programme Committee

Dr.Steven Furnell, University of Plymouth, Plymouth, United Kingdom
Dr. Paul Dowland, University of Plymouth, Plymouth, United Kingdom

Knowledge Management and E-Mobility

Jeanne Schreurs, Hasselt University, Diepenbeek, Belgium
Prof. Ricardo Chalmeta, Universidad Jaume I, Castellon, Spain

EUROSIS is a Division of ETI Bvba, The European Technology Institute, Torhoutsesteenweg 162, Box 4, B-8400 Ostend, Belgium

Printed in Belgium by Reproduct NV, Ghent, Belgium
Cover Design by Grafisch Bedrijf Lammaing, Ostend, Belgium

EUROSIS-ETI Publication

ISBN: 978-90-77381-4-65
EAN : 978-90-77381-4-65

PREFACE

The EUROMEDIA conference is the annual EUROSIS meeting aimed at exploring the latest in state-of-the art multimedia research, technology, management and art. As in previous years, the conference seeks to bring together researchers and practitioners in academia and industry, who are interested in exploring and exploiting new and multiple media to create new capabilities for human expression, communication, collaboration, and interaction. EUROMEDIA covers a broad range of topics, from multimedia computing: theory to practice, over underlying technologies to applications. The present event is no exception, providing an ideal forum for the presentation and exchange of research relating to the design and use of state-of-the-art multimedia and networked systems.

The EUROMEDIA 2009 conference, which is held at the Novotel Brugge Centrum from April 15-17, 2009, runs concurrently with the ECEC and FUBUTEC conferences, is structured around four main tracks (WEBTEC - which deals with web technology, COMTEC- which covers communications technology, MEDIATEC - which covers multimedia technology, and APTEC - which provides an overview of media-integration).

The EUROMEDIA 2009 conference brings together several researchers representing several fields related to web technology, multimedia technology, media-integration, communications technology and human computer interaction.

This book contains the full papers presented at EUROMEDIA 2009 conference that came from 10 different countries.

We would like to thank to Philippe Geril, whose continued dedication and hard work as the conference organiser has enabled us to maintain the standard expected of the EUROMEDIA 2009 conference, to EUROSIS for the opportunity to be involved in the organization of EUROMEDIA 2009, to Hasselt University for their support of the event, to all members of the Scientific Committee for their reviews and significant contribution for the high quality standards of EUROMEDIA 2009, to all sessions chairs for their effort for the smooth running of all scientific sessions of EUROMEDIA 2009, to our Keynote and Invited Lecturers and to all Authors for sharing their excellent works during EUROMEDIA 2009 and last but not least to all attendees that enrich and validate the purposes of EUROMEDIA 2009.

Prof. Jeanne Schreurs
Hasselt University
General EUROMEDIA 2009 Chair

CONTENTS

CONTENTS

SCIENTIFIC PROGRAMME

MULTIMODAL DATABASES

DATA MINING AND MARKETING APPROACH TO TRACK CUSTOMER MOVEMENTS

George Sammour , Jeanne Schreurs and Koen Vanhoof
[2]Hasselt University
Hasselt University, Diepenbeek Campus Agoralaan gebouw D, B3590
Diepenbeek, Belgium
E-mail: george.sammour@uhasselt.be, Jeanne.Schreurs@uhasselt.be,
koen.anhoof@uhasselt,be

KEYWORDS
Data Mining, email marketing, customer movement

ABSTRACT

Many business problems are of interest to both data mining and marketing academic communities, such as market segmentation, direct marketing, targeted marketing, personalization/customization, cross selling, discovering customer lifetime value and customer behaviour. Yet, these two disciplines have very different approaches to analyzing these problems. Data mining research incorporates methodologies from various research disciplines such as statistics, machine learning, database technology, optimization and pattern recognition, and hence has a richer pool of knowledge/model representation. On the other hand marketing research urges more theory-based analysis and its theories are often built upon statistics, economics, econometrics and other social sciences. In this paper we propose an approach to integrate and use a data mining methods for solving marketing problems specifically, the area of email marketing.

INTRODUCTION

Email Marketing

Although practitioners and academics have identified key success factors and key barriers to the development of an effective email campaign, few have attempted to apply existing theories and models. Similarly, although email marketing studies have been conducted either by online surveys, by in-depth interviews, by controlled experiments or by tracking behaviour patterns such as click-through links and the visiting patterns, few research have investigated the effects of email characteristics on consumer attitudes and behavioural intentions.
There are two types of research in email marketing. The first includes focus specifically at reducing spam from a wide range of perspectives. The second includes studies from the marketing literature that examine factors which affect and improve response rates, open rates and click rates for email marketing campaigns. The focus of this research will be situated in the second stream of email marketing research.
The context of this research falls in the second category of email marketing, which is improving response rate as we will analyze data of email campaigns sent to customers to increase response rate.
Another technique is permission marketing (Godin, 1999), which seeks permission in advance from consumers to send marketing communications. Consumers provide interested marketers with information about the types of advertising messages they would like to receive. The marketers then use this information to target advertisements and promotions. The aim is to initiate, sustain and develop a dialogue with customers, building trust and over time stimulating the levels of permission, making it a more valuable asset (Kent & Brandal, 2003). Permission marketing has three specific characteristics that set it apart from traditional direct marketing (Godin, 1999):

- Anticipation: Customers who permit their names to be included on direct-mail lists can anticipate receiving commercial messages.
- Personalization: The sending company can personalize those messages.
- Relevance: The messages will be more relevant to the customers' needs.

These characteristics are what allow marketers to cut through the clutter and speak to prospects as friends. This personalized, anticipated, frequent, and relevant communication has a greater impact than a random message displayed in a random place at a random moment.

With email marketing, using preferences stated by customers to select email content can be straightforward and based on common sense. However, there might be other customer-related factors, besides content matching stated preferences, which have an influence on the customer's open and click behaviour. For example, stated preferences or socio-economic information might reveal interests in non-stated preferences or might reveal hints about when to send out an email campaign. Using common sense to adapt email newsletter to such type of customer information will not guarantee exploiting its full potential. For this end, one needs to experiment how people with specific preferences or specific socio-economic background react to certain changes in the email campaign. However, most companies are reluctant to experiment with their email campaigns because they fear that the response rate will drop due to wrong experimenting. What we need is a methodology which allows experimenting with email campaigns yielding a high potential of increasing response rate levels while at the same time lowering the risks of detrimental effects due to

unsuccessful experimenting. In this article we propose such methodology.

Data Mining

"Data mining is the process of exploration and analysis, by automatic or semi-automatic means,of large quantities of data in order to discover meaningful patterns and results." (Berry & Linoff, 1997, 2000)
Some defining attributes are: a large set of data, an automatic analysis and being proacted over time.
For marketing, data mining is used to discover patterns and relationships in the data in order to help make better marketing decisions. Data mining can help spot sales trends, develop smarter marketing campaigns, and accurately predict customer loyalty. Specific uses of data mining include:

- Market segmentation
- Customer churn
- Fraud detection
- Direct marketing
- Interactive marketing
- Market basket analysis
- Trend analysis

Data mining tools predict behaviors and future trends, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Some of the tools used for data mining are:

- *Artificial neural networks* - Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- *Decision trees* - Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.
- *Rule induction* - The extraction of useful if-then rules from data based on statistical significance.
- *Genetic algorithms* - Optimization techniques based on the concepts of genetic combination, mutation, and natural selection.
- *Nearest neighbor* - A classification technique that classifies each record based on the records most similar to it in an historical database.

TRACKING CUSTOMER MOVEMENTS

By analyzing the tracks people make through their Web site, marketers will be able to optimize its design to realise their dream – maximizing sales. Information about customers and their purchasing habits will let companies initiate E-mail campaigns and other activities that result in sales. Good models of customers' preferences, needs, desires, and behaviors will let companies simulate the good personal relationship between businesses and their customers.
Visitor characteristics:

- Demographics
- Psychographics
- Technographics

While, web content information such as, media type, content category, URL as well as product information : SKU (stock-keeping unit, basically a product number), product category, color, size, price, margin, available quantities, promotion level, and so on. Item characteristics include

- *Visitor statistics:* accumulate when visitors (an individual that visits a Web site) interact with items, the Web site, or the company.
- *Visitor-item interactions:* this include purchase history, advertising history, and preference information.
- Click-stream information is a history of hyperlinks that a visitor has clicked on.
- *Link opportunities:* are hyperlinks that have been presented to a visitor.Visitor-site statistics include per-session characteristics, such as total time, pages viewed, revenue, and profit per session with a visitor.
- *Visitor-company information:* might contain total number of customer referrals from a visitor, total profit, total page views, number of visits per month, last visit, and brand measurements.
- *Brand associations:* are lists of positive or negative concepts a visitor associates with the brand, which can be measured by surveying visitors periodically.

It is also important to consider the information that we need to analyse and make them clear, when building a marketing web site, or designing an email marketing email campaign. Information can be prescribed as questions to be answered to help marketers need to know, we have translated into categories as shown in Table1.

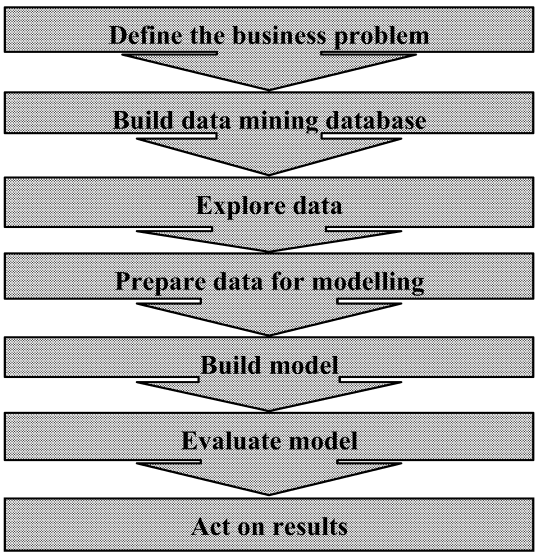
Table 1: Question nformation categories

What marketers ask?	What Marketers mean?
Who visited the web site?	Visitor ctegories (demographic or behavioral) sorted by visit frequency
Where did they heard about the website?	Ad compaigns or inbound hyperlinks sorted by visit frequency
What did they do?	Content category, for each visitor category, sorted by page view frequency
How did they use the site?	Traffic patterns next-click or previous-click from each page, sorted by frequency
How did they leave?	Exit pages, for each visitor category, sorted by visit category

Challenges of customer movements

The main goal of establishing a marketing website or starting an email marketing campaign is to maximize sales. The foundation of this goal is the log of customer accesses maintained by Web servers. A sequence of page hits might look something like this:
Page A => Page B => Page C => Page D => Page C => Page B => Page F => Page G. Or more explicitly:
Login => Register => Product Description => Purchase.

By analyzing customer paths through the data, vendors hope to personalize the interactions that customers and prospects have with them. Companies will customize the home page each customer sees, the responses to requests, and the recommendations of items to purchase.
To look at some special challenges of customer movements, let's examine the issues in the context of the data-mining process.
It's through data mining that companies can build the most effective models of their customers and prospects, Figure 1 reveals the process that is performed when applying data mining techniques to date collected from a marketing web site or an email marketing campaigns. As shown the first step is to define the business problem and the goal of the research to take place, note that this can be achieved by predefining and preparing the question information categorized in Table 1.



In the following subsections we will discuss each step in the process in more details.

Define the business problem

The typical goals in defining the business problem might include

- Improving the design of a Web site by identifying the paths people take to arrive at a purchase;
- Detecting problems such as pages that are never accessed;
- Suggesting strategies for increasing market basket size;
- Increasing the conversion rate (turning visitors into purchasers);
- Decrease products returned;
- Increase number of referred customers;
- Increase brand awareness;
- Increase retention rate (such as number of visitors that have returned within 30 days);
- Reduce clicks-to-close (average page views to accomplish a purchase or obtain desired information);

Building the data-mining database, exploring the data, and preparing it for modeling are the most time-consuming. For clickstream data, these tasks are particularly difficult, consuming 80% to 95% of a project's time and resources.
These are the key steps in building a data-mining database

- Integrate logs
- Remove extraneous items from log
- Identify users and sessions
- Complete paths
- Identify transactions
- Integrate with other data.

There are three approaches to identify sessions from Web access log data. 1. to use heuristics. IP addresses aren't enough to identify a customer because they're not unique to that person. Frequently, an IP address is assigned from a pool of addresses by an Internet service provider (America Online – Vienna, Va.). To identify a session, you can try a combination of IP address, browser type, and pages viewed.2. to embed session identification numbers in the URL. This works well as long as the customer doesn't visit another site during the session. If that happens, the session ID is lost upon return and the customer will appear as a new customer.3. to use cookies. A cookie is a text file placed on your computer that contains information about your session and what you did. Many customers don't like cookies, so they refuse to accept them or accept them only selectively. These surfers worry about being tracked or about having mysterious files residing in their computers.

Explore the data

Aggregations and distributions to quantify the following:

- How many people come to a particular Web site?
- Which sites refer the most visitors, and which sites refer the most visitors who buy something?
- How many visitors add something to a market basket?
- How many complete the purchase, and which searches failed ?
- What are the best-selling and worst-selling products?

Visualizations are a useful way to understand your data. By condensing information into a display, graphics let you quickly see how data is distributed, spot unusual values, or notice possible relationships among variables.•

Prepare data for modelling

Data transformation is the last step before building models. For example, in trying to predict who will be likely to respond to an offer, you may need to create new variables that are derived from your data. If you're working with existing customers, then RFM variables can be very good predictors.

- Recency - the number of days since the last purchase.
- Frequency - the number of purchases the last three months.

- Monetary - the total purchases in the last three months as well as the average order size over that period.

Build a model

Building the model depends on the type of research to be done on the data, there are two main types of data mining methods to build the model. First are collaborative filtering or association (*Discovery methods*), which can include product recommendations to customers based on previous purchases, the item being viewed, or the contents of a shopping cart. The drawback of these methods is that they are inaccurate (don't involve the testing phase of true predictive models), on the other hand, require much less information than more precise predictive models (as based solely on behaviors at the vendor site). Second, is *Predictive models,* which deal with factoring of information about characteristics and preferences of site visitors whose identity is known, the characteristics of these methods is that they are accurate and more customized prediction.

Evaluation of the model

It's important to evaluate models for accuracy and effectiveness. Effectiveness may be measured by such traditional economic metrics as profitability or return on investment.
However, these objective measures are useless if the model doesn't make sense.

Interpretation. Implemetation.

In Online marketing, there are two main classes of customer interaction. inbound - the customer comes to the site and outbound - the vendor goes to the customer, as in an E-mail promotion.
Inbound interactions require quick response to the various stages of the transaction. The relevant information, such as the identity of the customer and items in the shopping cart, must quickly be sent from the current transaction to the modeling engine, which determines the correct action and sends it back to the application. Outbound interactions are a bit more leisurely. To identify the targets of a campaign solicitation, the model can be applied in batch to the list of prospective recipients. And the actual effectiveness of the models must be compared with the reality, and if necessary the models and data modified as part of a continuous process of improvement.

Pitfalls and obstacles

Many decisions are made that may limit what can be discovered using DM, e.g.
- data warehouse attributes
- variables selected for analysis
- types of models considered
- observations selected

- Data are observational
- Observations are not rendomly selected
- Important variables may be unavailable
- Incorporating prior knowledge and avoiding „discovery of the obvious“
- Privacy issues
- Results may not be usable, interpretable, or actionable

APPLICATIONS OF DATAMINING

Targeting

Marketers use targeting to select the people receiving a fixed advertisement, to increase profit, brand recognition, or other measurable outcome. Targeting on the Web must account for different advertising ad space costs. Web sites with valuable visitors typically charge more for ad space. On sites where visitors register, advertisers can target on the basis of demographics. Some sites let you target ads on the basis of IP address. Data mining can help you select the targeting criteria for an ad campaign.
Web publications have a set of variables by which they can target advertisements. By performing a test ad using "run-of-site" (untargeted) ad space you can associate demographic variables with conversion. People "convert" when they accomplish the marketing goal, such as performing a click-through, purchase, registration, and so on. Data mining can identify the combination of criteria that maximizes the profit. For example, data mining might discover that targeting based on the logical expression*(java-consultant) or (software-engineer and purchasing-authority < 10,000)* will increase the click-through on a JavaBean banner ad.

Personalization

Marketers use personalization to select the advertisements to send to a person, to maximize some measurable outcome.
Bare in mind that personalization is the converse of targeting.
Personalization optimizes the advertisements that a person sees, raising revenue because the person sees more interesting stuff. Personalization can be used for external advertising.
Some personalization systems, such as Broadvision One-to-One, rely on the marketer to write rules for tailoring advertisements to visitors. These are "rules-based personalization systems." If you have historical information, you can buy data-mining tools from a third party to generate the rules. These systems are usually deployed in situations where there are limited products or services offered.
Other personalization systems, such as Andromedia LikeMinds, emphasize automatic realtime selection of items to be offered or suggested. Systems that use the idea that "people like you make good predictors for what you will do" are called "collaborative filters." These systems are usually deployed in situations where there are many items offered.

Knowledge Management.

These systems identifies and leverages patterns in natural language documents. A more specific term is "text analysis“. The first step is associating words and context with high-level concepts. This can be done in a directed way by training a system with documents that have been tagged by a human with the relevant concepts. The system then builds a

pattern matcher for each concept. When presented with a new document, the pattern matcher decides how strongly the document relates to the concept.•This approach can be used to sort incoming documents into predefined categories. Companies use this approach to build automatic site indices for visitors.
Knowledge management systems can be used to personalize online publications. Knowledge management systems can assist in creating automatic responses to help requests.

CONCLUSIONS

The use of the online market research methods is growing at the exponential pace. However, they will not replace traditional offline methods. Data mining, indeed, facilitates and supports market reserch by:

- Automated prediction of trends and behaviors: Data mining automates the process of finding predictive information in a large database.
- Automated discovery of previously unknown patterns: Data mining tools sweep through databases and identify previously hidden patterns.

Data mining is used to discover patterns and relationships in the data in order to help make better marketing decisions. Data mining can help spot sales trends, develop smarter marketing campaigns. Data mining techniques find predictive information that market experts may miss because it lies outside their expectations.
By tracking people through their Web site, marketers will be able to optimize its design to realise the ultimate goal – maximizing sales
Application of data mining techniques by many firms proves their usefulness, effectiveness and crusial meaning in market research and, consequenly, in performance of the whole economy.

REFERENCES

Ansari Asim, Mela Carl F. E-customization. J Mark Res 2003;40(2):131–46.

Fayyad, U.M, Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and Data Mining. AAA1 Press/The MIT Press, pp 1-34.

Godin, S. (1999). Permission marketing: Turning strangers into friends and friends into customers. New York: Simon Schuster.

Godin, Seth (2000), "Permission Marketing," Credit Union Executive, 41 (January), 42.

Godin, Seth (2000), "Permission Marketing," Credit Union Executive, 41 (January), 42.

Goldman Eric.ACoasean analysis of marketing.Wis Law Rev 2006;6(4):1152–221.

Goodman Joshua, Heckerman David, Rounthwaite Robert. Stopping spam. Sci Am 2005;292(4):42–5.

Gratton E. Dealing with unsolicited commercial emails: a global perspective. J Int Law 2004;7(12):3–13.

Interactive Prospect Targeting (IPT) Limited, 2006, The BIG BOOK of eMail Marketing, IPT Limited, ISBN 0-9552194-0-x.

J. Ross Quinlan , 1992, C4.5 programs for machine learning, Morgan Kaufmann Publishers, ISBN: 1-55860-238-0.

Mort G.S.; Drennan J. Mobile digital technology: Emerging issue for marketing. The Journal of Database Marketing, Volume 10, Number 1, September 2002, pp. 9-23(15)

BIOGRAPHY

GEORGE SAMMOUR is a PhD student in Business Informatics at the Transportation Research Institute — Hasselt University, Diepenbeek, Agoralaan gebouw D, 3590 Diepenbeek, Belgium.

JEANNE SCHREURS is Professor at the Hasselt University, Faculty of Applied Economics Campus Diepenbeek, Agoralaan gebouw D, 3590 Diepenbeek, Belgium.

KOEN VANHOOF is the vice dean of academic research for the Faculty of Applied Economics at Hasselt University, Campus Diepenbeek, Agoralaan gebouw D, 3590 Diepenbeek, Belgium.

DATABASE FOR A VISUAL LANGUAGE-BASED APPLICATION, Not Just a Collection of Image Files

Siska Fitrianie and Leon J.M. Rothkrantz
Man-Machine-Interaction, Delft University of Technology
Mekelweg 4 2628 CD Delft, the Netherlands
E-mail: {s.fitrianie, l.j.m.rothkrantz}@tudelft.nl

KEYWORDS
Visual language, usability, icon database, mobile application.

ABSTRACT
This paper reports our research in developing a database for our visual language-based application. The application offers inter-communication between people, where messages are created using spatial arrangements of visual symbols, i.e. icons. The current implementation is applied as a reporting tool in crisis situations using mobile communication devices i.e. PDAs and smart-phones. User experiments have been done to have more knowledge how to design good icons and a visual language-based interface. A corpus-based approach was used to have more insight of how humans express their concepts or ideas using this type of messages. By taking the display pipeline of current mobile devices and requirements for appropriate applications into account, approaches how to handle a large number of graphical data on mobile devices are discussed. These work result in guidelines that are used in developing the system database. This knowledge is referred as ontology – a representation of real world environment. To allow a widespread application, we also investigate the influence of context in developing the database.

INTRODUCTION
Visual language refers to the idea that communication occurs through visual symbols, i.e. icons, as opposed to verbal symbols or words. Sketches and images represent important tools for communication research and praxis (Singhal & Rattine-Flaherty, 2006). The visual language provides an alternative to the privileging of text and writing as a mode of comprehension and expression, especially to communicate about topics that are difficult to speak about. The icons represent objects and events in the world, for example fire, car, smoke. This representation supports faster interaction (Kjeldskov & Kolbe, 2002), reduces the ambiguity of the communicated information (Norman, 1993), and provides a language independent message construction (Perlovsky, 1999).

The needs of being able to access information anytime and anywhere makes mobile devices, PDAs and smart-phones, more popular due to its portability and facility for wireless connection. They are advancing with increasing numbers of features and traditional desktop applications. However, text input on such devices is still a bottle-neck (Karlson et al., 2006). Recent research has been done on adding multimodal capabilities, e.g. Comerford et al. (2001), Dusan et al. (2003), Wahlster (2006). However, the current technology makes speech input less suitable for mobile activities (Bousquet-Vernhettes, 2003)'Therefore, we aimed at a natural interaction style based on the strength of the Graphical User Interface using icons.

An icon is stored as a kind of image document, such as BMP, JEPG and GIF. A visual language interface naturally involves a large number of icons. In contrast to the text, handling images are more expensive regarding resources of the mobile device. Especially if a large number of graphics must be processed the system's limitations are quickly reached. The relatively small screen size is one of the major drawbacks if a large number of graphical contents should be displayed at once. This problem is also related to the provided screen resolution, which is also lower than in stationary gadgets. Together this leads to less graphical data which can be displayed. Moreover, an image is formed by many pixels which are discretely distributed, special techniques are required to present and manipulate a large number of graphical contents on mobile devices. The other bottleneck is lack of processing power. Due to the size of mobile devices, it is not possible to include hardware with similar performance as in stationary devices. This heavily affects the processing time, and therefore, usability of graphical data during presentation and interaction.

Despite the obvious applicability of icons to the design of a visual language interface, poorly designed icons can have disadvantages: (1) ambiguity occurs when an icon holds more than one meaning, (2) every user assigns a meaning to an icon depending on the individual's prior knowledge, (3) icons cannot always completely replace words, (4) it is costly to design new icons which is interpreted properly by many users and (5) displaying many icons at once can make the user confused. A careful design of icons, consistency, and respect for conventions are all essential.

To show in which circumstances the use of certain approaches leads to better results for the presentation of a large number of graphical and the user interaction of a visual language-based interface on mobile devices is the aim of this publication. Our consideration is device-oriented, user-oriented and system-oriented. This paper is structured as follows. Related work and our developed system are presented briefly in two sequential sections. We continue to discuss requirements on the use of specific media and interaction techniques that influence the performance. Then,

* The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024

user experiments in gaining knowledge how people recognize and recall icons on a visual language-based interface are described. The next is the design of our database, which is referred as ontology. Finally, we close our contribution with guidelines for an icon design, a visual language-based interface design and a large number of graphical presentations on mobile environments.

RELATED WORK

Recent attempts have been made to develop computer-based icon communication systems. This research mainly aimed at supporting people to communicate with each other despite not sharing a common language. Some systems are dedicated for speech impairment people (Basu, 2002; Barrow & Bakker, 2005). A message can be composed using an arrangement of icons or using combinations of symbols to compose new symbols with new meanings (Bliss, 1984; Housz, 1996). The arrangement can be realized in a rigid linear order (Yazdani & Mealing, 1995; Leemans, 2001) or a non linear order (Housz, 1996). Some of them produce text as the output of the input interpretation. Some systems are hard to learn, language specific, based on complex linguistic theories or using non-intuitive icons.

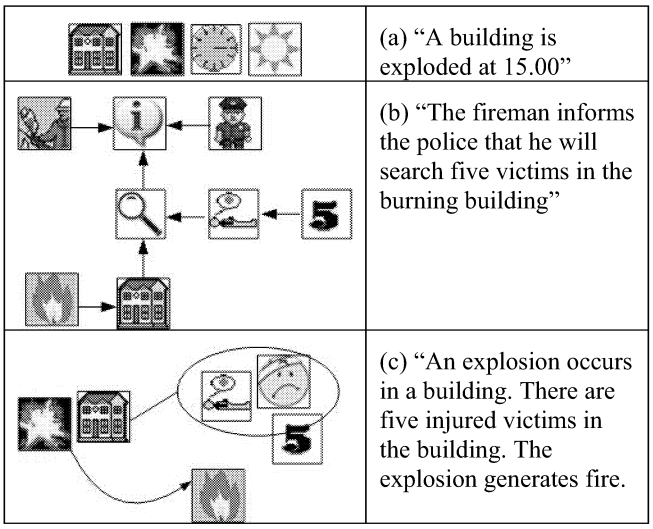


Figure 1 Examples of icon messages: (a) in a sequential order, (b) in a 2D configuration represented by arrows and (c) using lines, arrows and ellipses

Leemans (2001) developed an icon-based communication system based on the notion of simplified speech to emphasize simplicity and ease of use. Such research continued by the development of Lingua (Figure 1 (a) and (b) - Fitrianie & Rothkrantz, 2005; Fitrianie et al., 2006). Most systems mentioned here are based on linguistic theories, such as the Conceptual Dependency Theory (Schank, 1972) and Basic English (Ogden, 1930). An icon-based communication interface that allows for a free and natural way of creating a spatial arrangement of graphics symbols (i.e. icons, lines, arrows and shapes) to represent concepts or ideas has been developed (Figure 1 (c) - Fitrianie et al., 2008). The developed system provides drawing tools and predefined sets of icons. The interface creates a coherent and context dependent interpretation of the icon configuration and results scenarios as feedback to the user input.

To date, we have not yet found a standard icon database that can satisfy all requirements of a visual language-based application. Most applications develop their own icon database that is compatible for their visual language. An icon sentence is composed by a spatial arrangement of icons (Chang et al., 1994). Each icon represents meanings. An individual icon can be interpreted by its perceivable form (syntax), by the relation between its form and its meaning (semantics), and by its usage (pragmatics) (Chandler, 2001). To be able to express a certain meaning, like a spoken sentence, an icon sentence can be formed by a number of grammatical units of icons like nouns, pronouns, verbs, adjectives, and prepositions. Zlango, an icon-based language for mobile messages, chat and e-mail, offers more than 200 icons including different grammatical units (Arlington, 2006). The language conveys simple everyday sentences and a direct translation of each icon in a sentence (Figure 2(b)).

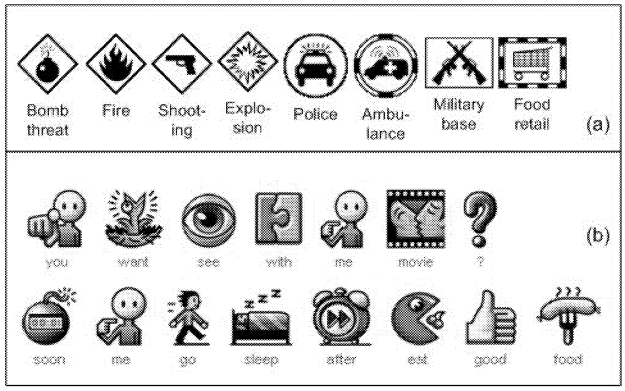


Figure 2 (a) Examples of a set of icons from HSWG (2003) and (b) two icon sentences from Zlango (2006)

As pictorial signs, icons can be analyzed using semiotics, which concerns how signs obtain their meaning and how they convey them (Chandler 2001). According to Peirce (1955), the degree to which the user interaction exists depends on how close the interpretation in the user's mind is to the object of an icon, e.g. (1) iconic signs occur when the representation relates to the object through a resemblance, (2) indexical signs occur when the representation relates to the object through causation, and (3) the representation symbolic signs relates to the object purely through convention. Since this interpretation is a subjective matter, in particular, the approach shows that designers should include the interpretations of different viewers on the evaluation of their icon designs (Horton, 1994). During icon design process, available guidelines and standards can be followed, e.g. ISO/IEC 2000, Horton (1994) and Schneiderman (1992). In the field of crisis management, research on guidelines and standards for the design of hazard and emergency maps and icons, including those of US military and NATO, has been done by Dymon (2003). Based on this research, standard map symbols are promoted by the U.S. Government on a national basis (HSWG, 2003 - Figure 2(a)) for sharing information during emergency situations by emergency managers and people responding to disasters.

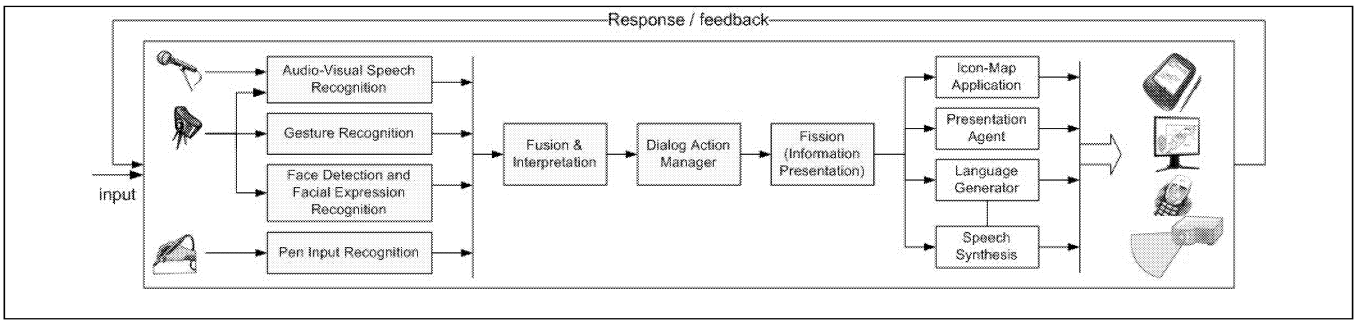


Figure 3 The developed framework architecture for single user

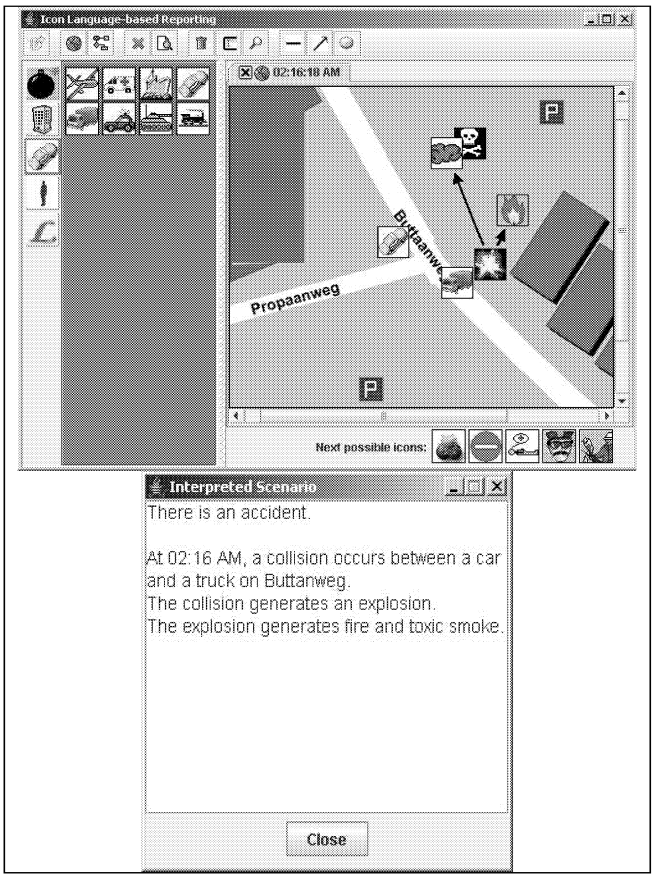


Figure 4 An example of visual-language based application for reporting situation

OVERVIEW: A VISUAL LANGUAGE-BASED APPLICATION

The introduction of novel information and communication technology in the crisis management domain can help to provide more detailed and accurate situation overviews that are current and shared amongst all management levels. Towards this goal, we have developed a framework that allows the rapid construction and evaluation of multimodal human-computer interaction (HCI) systems (Fitrianie et al., 2007). The development aims at module integration that is independent of the availability of modalities. Currently, we developed a project demonstrator system for reporting observations, which is able to collect many small but up to date observation reports, interpret them automatically and form a global view about the reported events. The framework includes input recognition modules of different modalities, such as text, speech, visual language, gesture, pen writing and drawing, and facial expression (Figure 3). The output combines text, synthesized speech, visual language and control of the underlying user interface. A presentation agent that is able to generate appropriate speech with intonation and facial behaviors is also utilized.

We designed a map interface to support people with geospatial information (Figure 4 - Fitrianie et al., 2008). It provides drawing tools for a free and natural way of describing a crisis situation using predefined icons, lines, arrows, and shapes. The interface provides icon-strings (Figure 1), text and photos input for non-spatial information. A coherent and context dependent interpretation and textual crisis scenario of the icon configuration are constructed as feedback to the user input.

The framework is designed to support various roles within the crisis management, including rescue workers and civilians in the field, which mostly work with mobile devices and control room operators, which mostly work with stationary devices. A centralized Fusion Manager integrates every newly reported situation from all users and adapts the world view accordingly then sent it back to the network.

REQUIREMENTS FROM DEVICE PERSPECTIVE

In displaying graphics, data passes a number of steps. The first is loading the graphical content to memory, which is influenced mainly by properties like file size and format. Since the content is often encoded, loading means file reading and decoding in memory. More detailed statements can be derived about affected properties of the device. The decoding step transfers the content of a file to an internal memory representation (IMR), which form the basis for later display. The IMR is mainly influenced by the properties image size and precision and the actual image content. The display step shows either the whole or portion of the IMR on display. Hardware-dependent and currently compulsory is the use of a discrete raster display with a certain screen resolution. A transformation function is used to map graphical content from a logical coordinate system to the physical display coordinate system. Kind and properties of the transformation function are determined by the graphical content, user interactions and a number of other points.

Vector and raster data are based on completely different ways to describe graphical content. Raster graphics are used in areas where a content description by geometrical objects is difficult or even impossible, e.g. in digital imagery. The image content of raster graphics is described by discrete image points (pixels) on a regular 2D-grid of certain image size and precision. A common sizes of icons in most GUI-based software are 16×16 pixel and 32×32 pixel, high resolution program can support the maximize size up to 64×64 pixel. The ideal case is under different resolutions should have a set of corresponding big icons and small icons. Every pixel is independent from each other regarding its color, which causes a quite large file size. Thus, raster data is often stored in compressed representation, e.g. in GIF- or JPEG-format, only sometimes uncompressed, e.g. in BMP-format. Small compressed files need generally more processing power for decoding, but are faster to read. However, the structure of the resulting IMR, mostly a bitmap and no conversion is necessary to map the IMR to display since both are based on the raster approach. Modifications (e.g. zooming) might be imposed by the transformation function, which influences the presentation quality.

Vector graphics use geometric primitives and their attributes to describe image content. Such primitives are for instance points, lines, rectangles, or circles, with attributes like fill- and stroke-color, or stroke-width. Graphical content described by vector primitives is in general smaller than raster data depending on the number and complexity of primitives. The most obvious approach is to store a description of the vector primitives as IMR, and to render them directly to screen at display time (direct drawing).
For modification, a transformation matrix is preliminarily applied to primitives. Since no information gets lost during the transformation this delivers very good visual results, but costly processing power. It can be useful to render the vector graphic after decoding to an IMR-bitmap and to discard the primitives (indirect drawing).

Research of Rosenbaum et al. (2004) found that in the case of simple graphics like icons (that can be described by around 100 vector primitives), vector graphics are best to reduce constraints caused by limitations of the current mobile hardware. Furthermore, two main requirements for exploring a large number of graphics on mobile devices should be fulfilled (Rosenbaum & Tominski, 2003): (1) high presentation quality and (2) short presentation and update time. Since interaction is essential, performance is influenced by the used interaction technique. We constrain our considerations to the elementary zoom and pan approach. Interaction further requires the distinction in presentation, which spread from loading until displaying the content, and update time, which considers the time to display the IMR.

REQUIREMENTS FROM USER PERSPECTIVE

User experiments have been conducted to answer three questions: (1) which icon can represent a certain concept; (2) how to group icons so that they are easy to find and (3) how to express a sentence by icons. Eight people took part in the experiments and one person as a pilot. There was no assumption on distinction of gender or cultural differences.
The experiments consisted of four parts. The first is an icon design test based on Horton's icon evaluation. The participants were asked to recognize fifty icons. These icons were selected to include different concepts and varying levels of the possibility of ambiguity. We matched the answers against the actual answers given by the designer. For not-understand and ambiguous icons, we analyzed them for future design process. In the second part, the participants were asked to divide a hundred icons into groups of concepts. They may use some or all provided icons. We matched the resulted grouping to one another. In the third part, we gave the participants six different pictures and asked to arrange icon messages to represent the situation in the pictures. If they could not find any suitable icon, the participants might suggest their own opinion by drawing a new design of an icon. We translated each answer based on the icons' grammatical units. The final part is an interview mainly about their experiences and problems during the experiments. We also asked them to describe their process to accomplish all tasks in every section.

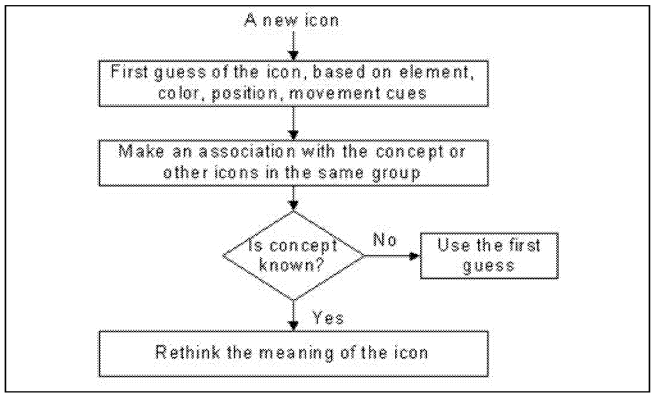


Figure 5 The flowchart of the thinking process in recognizing an icon on a concept

In recognizing a new icon, the participants made a guess about the individual icon based on the icon elements, the color, the position of each element, and movement cues (Figure 5). They would try to relate the icon's meaning with the provided concept or with other icons in the same group. If they did not understand the concept, they would abandon it and use the most probable initial guess. The experiment results indicated that iconic and indexical icons were easier to be recognized than symbolic icons. In the first part of experiments from twelve only one iconic icon cannot be recognized by our participants. Around 86% indexical icons were well recognized. However, from twenty-four symbolic icons only 50% were well recognized. Most of these icons represented abstract objects, prepositions, time, and verbs. Verb, adjective and preposition icons are recognized due to some cues elements, such as lines, arrows, words, and gestures. For example:  "go",  "up", "zzz." for  "sleep",  "pay" and  "bad". Simple

designed icons are easier and quicker to be recognized, for example - for angry instead of .
The participants could recognize popular icons easily, e.g. a scissor for "cut", a magnifying glass for "search", and a fork and a knife for "restaurant". However, 75% of our participants misrecognized "open" since it represents "exit" in many windows-based interfaces. Moreover, icons that have similar meaning can confuse users, e.g. the icon "not" with "stop".

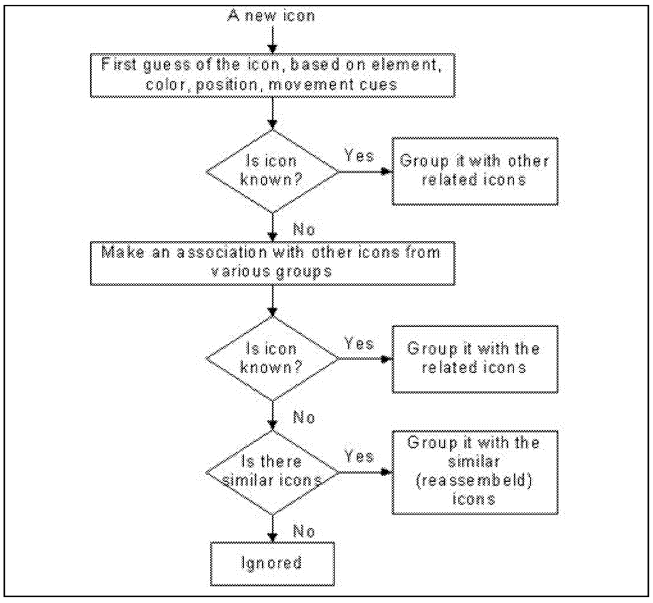


Figure 6 The flowchart of the thinking process in grouping an icon around a concept

In grouping an icon into a concept, the participants also made an initial guess (Figure 6). If the icon's meaning were still unknown, they would try to associate it with other icons. Most of the time, this technique could reveal the meaning of the icon. Otherwise, the participants would try to find the icons that have similar elements and group the icon with these icons without knowing the meaning. If there were not any similar icons, the icon would be ignored. All participants grouped the icons mainly based on categories, for example:

- building (house, museum, shop).
- notation (not, question, exclamation).
- adjective (good, bad, danger).
- emotion (surprise, happy).

Only one participant tried to express every word in creating an icon sentence. Other participants expressed only important messages by two or three icons (Figure 7). We found that all participants never used any icon for question words (e.g. "where", "what", "when") and preposition (e.g. "at", "in"). The adjective icons would be used only if the meaning were important to construct the message. We found that it was preferable to use an icon with adjective meaning, e.g. "an empty cup". The verb icons would be used only if their meaning could clear the meaning of the sentence. For example, without any verb icon, " + " would mean: "it is not Japan" or "I am not Japanese".

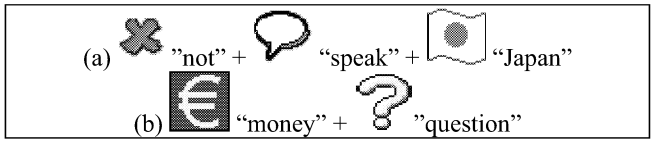


Figure 7 Two icon sentences (a) "I do not speak Japanese" and (b) "How much money do I have to pay?"

REQUIREMENTS FROM SYSTEM PERSPECTIVE

An icon sentence with more than just one icon can still be represented by the same icons, but it turn out to be difficult to determine the meaning. The meaning of individual icons represents a word or a phrase created according to the metaphors appropriate for the context of this type of languages. The meaning of the sentence depends on the context used to characterize the situation that is conveyed. The only thing that can be automatically derived from the semantics of the icons in an icon sentence is a fixed word or phrase belonging to these icons, e.g. "not", "speak", "Japan". The problem is that individual icons provide only a portion of the semantic of the sentence. The sentence meaning is derived as a result of the combination of these icons, and cannot be detected without a global semantic analysis of the sentence. It is unlikely that constructing an icon message with a one to one matching of icons to words would be appropriate. Humans, on the other hand, have innate ability to see the structure underlying a string of symbols that determine the possible shapes of human languages.

Knowing the message context can assist the system in the most appropriate interpretation of user icon messages. Furthermore, on a visual language-based interface, users will interact with a large number of icons. Grouping related icons based on their context is the only way to hint where an icon can be found because the meaning of icons almost exists in the context of other icons (Horton 1994). Moreover, this way can support a compact interface.
To "understand" the meaning conveyed on user messages, the knowledge must be modeled as the structured record of the user's world. The knowledge serves as information storage about entities, processes and events that are relevant to the interaction between a user and an application. The modeling of the underlying world models constitutes a major challenge: the information structures should share a common semantics and able to fulfill the context aware requirements. To represent a logical/abstract view about the world, ontology is selected for this type representation. Ontology forms a formal explicit description of concepts in a certain domain, properties of each concept, and restriction on slots.

Therefore, we are able to (Noy & McGuinness, 2001): (1) share common understanding of the structure of information among different users, (2) enable reuse and couple of domain knowledge in different contexts, (3) make domain assumptions explicit, (4) separate domain knowledge from the operational knowledge, and (5) analyze the knowledge.

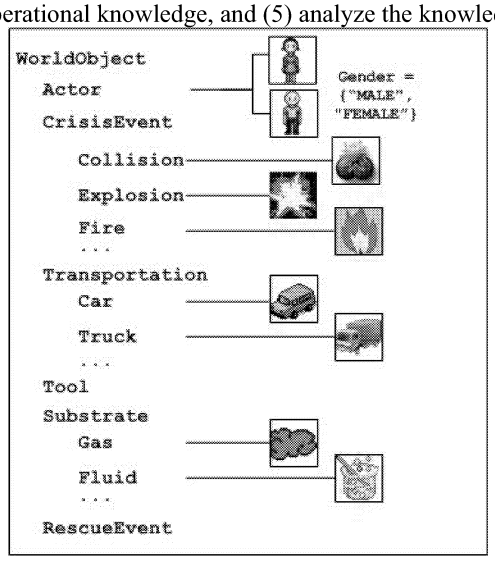


Figure 8 The WorldObject Taxonomy; the icons are the instances e.g. the icon "victim" is a subclass of an Actor

We develop the world model consisting of two geo-referenced contexts: (1) a chain of temporal specific events and a group of dynamic objects in action at a certain location in the world and (2) the geographical information concerning the crisis location, such as buildings, streets and parcels. The world model has direct links to the visual symbols on the user interface (Figure 8). Each concept represents an object, action or event in the world. It contains specific information of the event and object including its current status (e.g. living condition and dynamic spatial state), temporal information (e.g. frequency and time point), spatial information (e.g. current location, origin, destination and path) and relations among concepts (e.g. result, cause and contain).

An agent-based architecture for constructing coherence and context dependent interpretation of an icon message has been developed (Fitrianie et al., 2008 – Figure 9). The message interpretation is derived from purposive behavior emerging from interaction between multiple concepts belong to the arrangements of icons on the user's workspace. Agents are created to continually look for interesting properties and dynamically build relationships among these concepts and message structures. The context awareness can emerge as adaptations to the outcomes of interactions with the properties of each concept defined by the ontology, in pursuance of minimally specific crisis scenarios, which are defined by scripts. A script represents the chain of events that identify a possible crisis scenario (Schank & Abelson, 1977).

In information presentation, the ontology is used in the spoken language generation and visual language display (Fitrianie et al., 2008). The template-based *language generation* works concept-name recognition and substitution of the property-names by their values controlled modified-AIML format. AIML is an extended-XML script that provides specifications for a dialogue system (Wallace, 2003). The *visual language display* provides a representation of a map of a crisis area. Since some concepts have a direct link to the icons, the display of these concepts is considered as one-to-one mapping to their correspond icon on a certain location on a map.

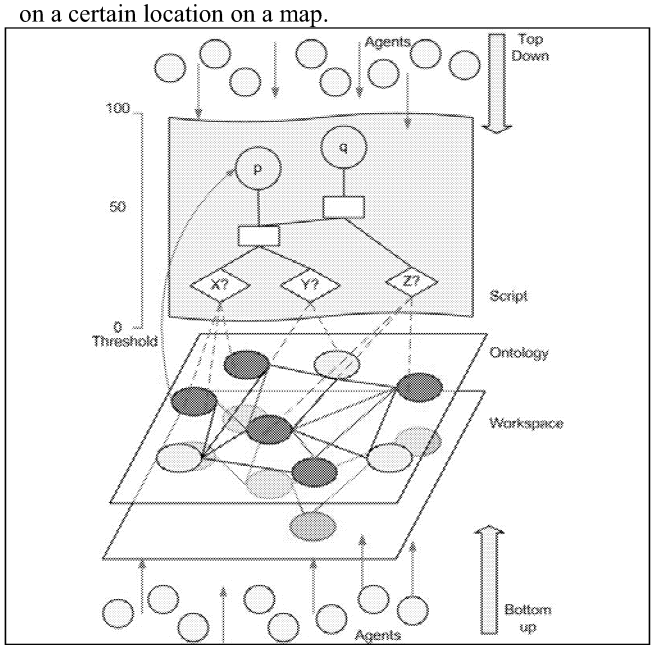


Figure 9 Agent-based icon message interpretation

CONCLUSION

In this paper, we have reviewed the display pipeline in mobile devices together with important properties of vector and raster data. User experiments on how users recognize and recall an icon among other icons and how users create icon messages have been performed. Demands of database structures for an input interpretation and appropriate information presentation have been stated. The research results can be summarized to the following guidelines.

From the mobile devices' point of view, graphics loading time depends strongly on the used file format. Loading icons (for up to about 100 primitives) is fast if they are stored as vector data. Although rendering vector primitives directly to display is generally slow and depends on primitive number and complexity, using an IMR-bitmap indirectly achieves faster update times. Moreover, scaling vector graphics by integer matrix multiplications is fast and achieves high quality. As computational power of future mobile devices increases, this broadens the application of vector graphics.

From the icon and visual language interface design, using popularly used icons is more usable, e.g. road sign icons, software interface icons and institution icons. However, their meaning should not be changed. The users can better recognize iconic and indexical icons than symbolic icons. Movement visual cues can be used for verb, proposition and adjective icons. Simple image with less number of elements is easier to be recognized. Moreover, an understandable concept or group of icons leads to understandable icons, i.e. based on their category like food, color, time. It is advised to

group indexical-verb icons apart from iconic icons. Furthermore, the translation of any user selection should be provided directly, the users can learn the icons on trials. Dealing with a large database of icons, a visual language-based interface should support a faster interaction, e.g. utilizing a search engine, an icon prediction, a syntax checker and distinctive cues of syntactically plausible icons. In this ways, the interface can offer reliable interpretation of inputted icon messages since mobile users cannot devote their full attention to operate the application.

From the system's point of view, context plays important role in interpreting user icon messages and presenting information. Without making the context, misinterpretation occurs. The context is resulted from the process of a cognitive awareness of the current situation and the state of the user. We approach this by modeling the system's knowledge as the structured record of the user world using ontology. The ontology is defined as a body of knowledge about the user, task and environment. It allows us to structure and manage different contexts in certain level hierarchies while developing the database. This makes a widespread application possible to different domains. In a long-term usage, it is advisable to allow users adding and editing vocabulary into the database. However, its operation should be designed in a way that the consistency of the icons database is still maintained.

REFERENCES

Basu A., Sankar S., Chakraborty K., Bhattarcharya S., Choudhury M. and Patel R. (2002) Vernacula educational communication tool for the people with multiple disabilities. *Development by Design Conference*. Bangalore.

Bevelier D., Corina D.P. and Neville H.J. (1998) Brain and language: a perspective from sign language. *Neuron*, Cell Press, 21:275-278.

Bliss C. K. (1984) *The blissymbols picture book*. Semantography Press, Sidney.

Bousquet-Vernhettes C., Privat R., and Vigouroux N. (2003), Error Handling in Spoken Dialogue Systems: Toward Corrective Dialogue, *Proc. Of ISCA'03*, USA.

Champoux B., Fujisawa K., Inoue T. and Iwadate Y. (2000) Transmitting visual information: icons become words. *IEEE – Information Visualization*, 244-249.

Chandler D. (2001) Semiotics: the Basic. Routledge.

Chang S.K., Polese G., Orefice S. and Tucci M. (1994), A Methodology and Interactive Environment for Icon Language Design, Int. Journal of Human Computer Studies. 41: 683-716.

Comerford L., Frank D., Gopalakrishnan P., Gopnanth R., and Sedivy J. (2001), The IBM Personal Speech Assistant, *Proc. of the ICASSP 2001,* USA.

Dymon U.J. 92003), An Analysis of Emergency Map Symbology, Int. Journal of Emergency Management. 1(3): 227-237.

Fitrianie S. and Rothkrantz L.J.M. (2005) Communication in crisis situations using icon language. *Proc. of IEEE ICME 2005*, the Netherlands, 1370-1373.

Fitrianie S., Datcu D. and Rothkrantz L.J.M. (2006) Constructing knowledge of the world in crisis situations using visual language. *Proc of. IEEE SMC 2006*, Taiwan, 121-126.

Fitrianie S., Poppe R., Bui T.H., Chitu A.G., Datcu D., Dor R., Hofs D.H.W., Wiggers P., Willems D.J.M., Poel M., Rothkrantz L.J.M., Vuurpijl L.G. and Zwiers J. (2007) A Multimodal Human-Computer Interaction Framework for Research into Crisis Management, *Proc. of ISCRAM 2007*, 149-158.

Fitrianie S., Yang Z., and Rothkrantz L. J. M. (2008) Developing concept-based user interface using icons for reporting observations. *Proc of ISCRAM 2008*, USA.

Fitrianie S., Tatomir I. and Rothkrantz L.J.M. (2008) A context aware and user tailored multimodal information generation in a multimodal HCI framework, E*uromedia 2008*, , Eurosis, Ghent, 95-103.

Horton, W. K. (1994) *The ICON book: visual symbols for computer systems and documentation*. John Wiley and Sons, Inc., New York, NY, USA.

Housz T. I. (1994–1996) *The Elephant's Memory*. http://www.khm.de/~timot.

HSWG - Homeland Security Working Group (2003), Symbology Reference. http://www.fgdc.gov/HSWG/index.html.

ISO/IEC (2000), International Standard, Information Technology – User System Interfaces and Symbols – Icon Symbols and Functions, 1st Edition, ISO/IEC 11581-1:2000(E) to ISO/IEC 11581-6:2000(E).

Leemans N.E.M.P. (2002) *VIL: A Visual Inter Lingua*. Dissertation, Worcester Polytechnic Institute, USA.

Littlejohn, S. W. (2001) *Theories of Human Communication* - with InfoTrac, Wadsworth Series in Speech Communication, Wadsworth Publishing.

Karlson A., Bederson B. and Contreras-Vidal J. 2006. Understanding Single Handed Use of Handheld Devices. Lumsden Jo (Ed.), Handbook of Research on User Interface Design and Evaluation for Mobile Technology, in press.

Noy N. and McGuinness D.L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*, Technical Report KSL-01-05, Knowledge Systems Laboratory, Stanford University.

Rosenbaum R. and Tominski Ch. (2003) Pixels vs. Vectors: Presentation of large images on mobile devices. *Proc. of IMC'03*, Germany, 2003.

Rosenbaum R., Tominski Ch. And Schumann H. (2004) Presenting Large Graphical Contents on Mobile Devices - Performance Issues. *Proc. of IRMA 2004*, USA.

Schank R. (1972) Conceptual Dependency - a Theory of Natural Language Understanding. *Cognitive Psychology*, 3, 552–631.

Schank R. and Abelson R. (1977) *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Erlbaum.

Schneiderman B. (1992), Designing the User Interface: Strategies for Effective Human-Computer Interaction, 2nd Edition, Addison-Wesley Publishing, 1992.

Singhal A. and Rattine-Flaherty E. (2006) Pencils and Photos as Tools of Communicative Research and Praxis. *International Communication Gazette*, 68(4):313-330.

Wahlster, W. (2006) Dialogue systems go multimodal: The Smartkom experience. *In SmartKom: Foundations of Multimodal Dialogue Systems*. Springer, Heidelberg.

Wallace, R. S. (2003) *The elements of AIML style*, A.L.I.C.E. Artificial Intelligence Foundation, inc.

Yazdani M. and Mealing S. (1995) Communicating through Pictures. *Artificial Intelligent Review*, 9(2-3): 205–213.

Knowledge Mapping and Modelling -- Based on RIA Multimedia information retrieval system

Chyi-Wen Hwang
Lan-Yang Institute of Technology, Taiwan
Institute of Education, University of London
welishwang@yahoo.com.tw

Abstract

Rich Internet Applications (RIA) like Ajax and Flash are now being used to provide a number of enhancements to Web applications;

In this paper, the researcher using flash skill to build up the Knowledge Mapping and Modelling --- Based on RIA Multimedia information retrieval system, to connect the Concept Map (CM) + Semantic Structure (SS) and the Knowledge on Demand (KOD) service. The aim of this empirical test was to investigating the utility and usability of this RIA Knowledge Mapping and Modelling will enhancement the User Interface responsiveness? And will improve User Interface?

Furthermore, based on the segment markets theory in the Marketing model, to propose a User Interface (UI) classification strategy, and formulate a set of RIA Mapping and Modelling design principles for further UI strategy & e-learning resources.

This research finding:

(1) With the help of RIA Knowledge Mapping and Modelling retrieval system, within just a single screen, Web applications can support a variety of complex user interactivity, and allows users to interact in real-time.

(2) Other than better user interactivity, RIA Knowledge Mapping and Modelling retrieval system --- allow a variety of option, also revolutionized the usage of video, animation, sounds and graphics in improving the user interface within the limited network bandwidth.

(3) Irrespective of whether the simple declarative knowledge or the complex declarative knowledge model is used, this RIA Knowledge Mapping and Modelling of "CM + SS + KOD retrieval system" has a better cognition effect than the "Non CM + SS + KOD retrieval system". However, for the"No web design experience user", this RIA Knowledge Mapping and Modelling does not have an obvious cognition effect.

(4) The essential of web classification design---: Different groups of user have a diversity of preference needs and different cognitive styles in this RIA Knowledge Mapping and Modelling retrieval system

Keyword: Knowledge Mapping and Modelling; RIA; Multimedia information retrieval; Concept Map (CM); Semantic Structure (SS); Knowledge on Demand (KOD); knowledge construction.

1. Introduction

This research is based on the Information Architecture design in Human Computer Interface (HCI) theory; and the Interactive; Sensorial design in User Interface (UI) theory, to explore the utility and usability of RIA Knowledge Mapping and Modelling retrieval system in adaptive Web learning environment.

Moreover, in adoption of Knowledge Space Theory, Dietrich Albert (1997) points out: the knowledge space provides a formal model for representing students' knowledge and describing the structure of domain knowledge. In this researcher's opinion, the RIA Knowledge Mapping and Modelling retrieval system with Concept Map (CM); Semantic Structure (SS) and Knowledge on Demand (KOD), this kind of integrated with the multimedia metadata and self-navigating format not only enhances the information communication process, but also fosters the user's cognition construction.

2. Knowledge Mapping and Modelling construction-- Based on RIA Multimedia (CM+SS+KOD) information retrieval system

This RIA Knowledge Mapping and Modelling retrieval system consists of three fundamental parts: "visual concept perception", "hierarchical structure representation", "knowledge organization and adaptive memory".

Jonassen, Beissner & Yacci, (1993) point out that "Structural knowledge" theory, can be used to present the relationships of knowledge content in a spatial way; with a different method to express the knowledge structure, the different types of organizing techniques, can promote the different cognition processes.

Table 1: Beissner, Jonassen & Yacci (1993): Cognition process and strategy

Method		Cognition process	Cognition effect	Cognition performance
Hierarchical	Semantic map	Analysis Organization Classification Elaboration Interpretation	Recall Transfer	Memory Exercise
Non-hierarchical	Concept Maps	Integration: Correlation Classification Establishment metaphor	Deduction	Searching

2.1. Retrievable & reusable system:

The visualized RIA Multimedia (CM+SS+KOD) information modelling is one type of knowledge retrieval system, for organizing and representing knowledge with the semantic content. It is related to multi-modal information, like the Hypermedia Streaming, VOD, MOD technology… etc. and can let the user chooses their options (knowledge) from the "popularized" to the " specialized", to reduce the searching procedures and repeat the cognition memory.

The RIA Multimedia (CM+SS+KOD) information retrieval system with adaptive & re-usable characteristics can increase the efficiency and effectiveness of query routing during the different metadata contents, and provide convenience to the user making their personal learning during the applications processes easier.

This empirical test model of the "RIA Multimedia (CM+SS+KOD) information retrieval system" can be divided into four parts of Layers: "Multimedia objects Layer"; "Subject and specimen multimedia archives Layer"; "Knowledge elements Layer" and "Classified browsing construction Layer".

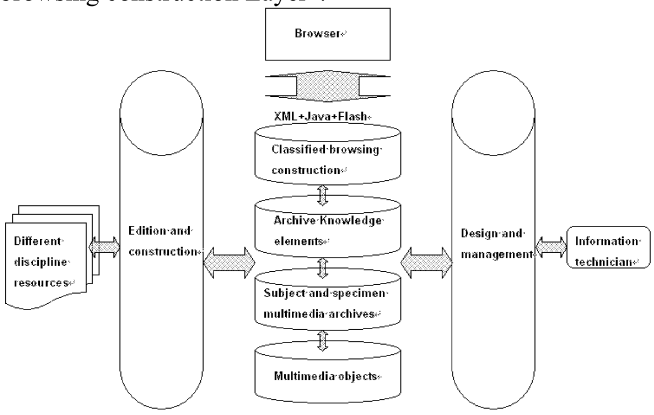


Figure 1: The System structure and processing pattern[1]

Moreover, the information metadata construction and retrieval process in this system is starting from "Data collection & classification" -- "Digital and visual processing" -- "Build up the Object Databank" -- "Data Combination and XML Translator" -- "Provide the Index Search; Full Text Search; Image Search; Hyper Link" and "Finally present on the Browser".

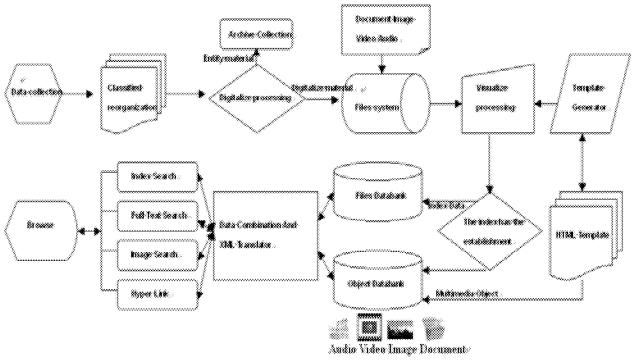


Figure 2: Multimedia metadata construction and KOD retrieval design

3. Research methodology

The researcher develop a RIA Mapping and Modelling in this research and aims at the Botany category for this empirical test sample (the test contents included simple declarative knowledge and the some complex terminology declarative knowledge). Firstly, the researcher designed a "No retrieval system" edition to compare with the "RIA

[1] Figure1 and Figure 2 was excerpt from the "Nation Digital Archives Program, Taiwan. http://www.ndap.org.tw/2_techreport/index.php?pid=212

Knowledge Mapping and Modelling of CM + SS + KOD retrieval system", further to explored what kinds of design strategies can motivate the user's cognitive comprehension; recall ability and learning retention. The researcher used random sampling and random clusters to choose 100 testees from the general public (for each experimental edition respectively), the testee's ages ranged from 18 to 65 years old.

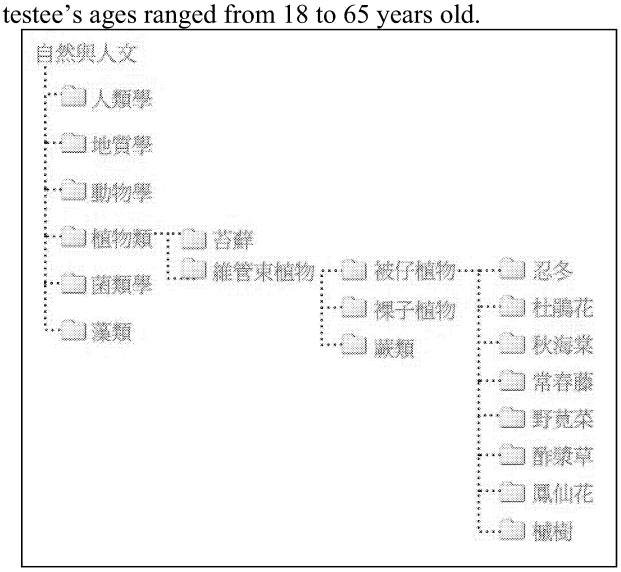


Figure 3: No retrieval system (Chinese edition)

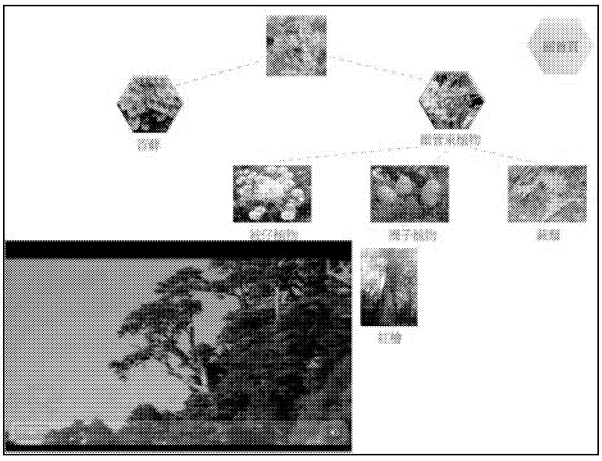


Figure 4: "RIA Knowledge Mapping and Modelling of "CM + SS + KOD retrieval system" (Chinese edition)

Secondly, as Duffy, Lowyck & Jonassen (1993) expressed: The different knowledge categories should have different knowledge structures; design methods and learning strategies. The researcher designed a questionnaire --- "RIA Knowledge Mapping and Modelling of CM + SS + KOD retrieval system", and according to the empirical test results to divided the testers into three groups: "Novice"; "Experienced"; "Expert".

This rating standard is adopted from Novak & Gowin (1984) and is used to analyze the different user's ability to construct the correct relationships and knowledge structure of Concept Maps & Semantic Structures. (As in the Figure 5 & Figure 6 below) If the user makes a proposition = 1 point; makes a Hierarchy = 5 points; makes a Cross-link = 10 points; makes an Example = 1 point. If the total score <10 = the testee belongs to the "Novice" group; if the total score>=10 and <50 =the testee belongs to the "Experienced" group; if the total score>=50 = the testee belongs to the "Expert" group.

Moreover, in order further to understand the user's learning attitudes and cognition results to this RIA Knowledge Mapping and Modelling retrieval system, The researcher also conducted in-depth interviews, to explore the different groups of user's diversity demands on the retrieval system and design strategies.

The researcher also uses random sampling to choose 118 testees and 21interviewees, who had learned web site design in previously. The testee's ages ranged from 18 to 65 years old.

Figure 5: The complete diagram of CM

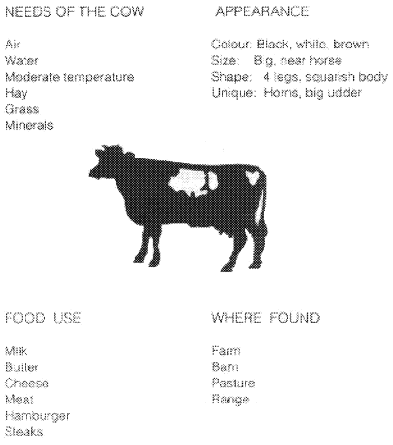


Figure 6: The sample question of SS

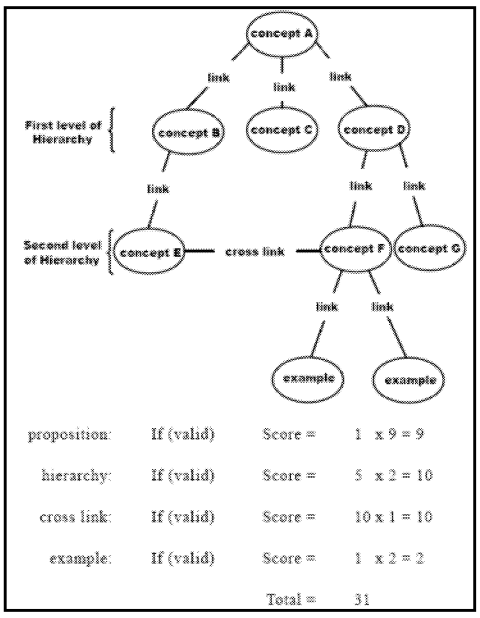


Figure 7: Novak & Gowin rating standard

4. Research finding (1): Does the user need a RIA Knowledge Mapping and Modelling retrieval system?

The data was analyzed comparing the pre-test & post-test results to assess the validity of the "No retrieval system" against the "RIA Knowledge Mapping and Modelling of CM + SS + KOD retrieval system" and using SPSS to analyze the Paired-Samples; T-Test (test-retest reliability).

In the "No retrieval system" edition, the validity of Correlation = 0.906, P = 0.591>0.05; in the other case, the "Have CM + SS + KOD retrieval system", the validity of Correlation = 0 .875, P=0.591>0.05. Cronbach's Alpha = 0.890.

Comparing the results from these two editions indicated that, for the "No web design experience groups", Irrespective of whether the "RIA Knowledge Mapping and Modelling retrieval system"" or the "No retrieval system", there was no obvious difference and have no urgent need than the "Have learned Web Site Design Group".

Table 2: Pre-test & Post-test result

	Correct answer rate of "Have CM+SS+KOD" retrieval system	Correct answer rate of "No retrieval system"
No web design experience groups in **Pre-test**	69%	66%
No web design experience groups in **Post-test**	79%	70%
Have learned Web Site Design Group in **Pre-test**	81%	68%

Comparing the pre-test result from six months before, 24 testees of the "No web design experience groups ", took a professional computer curriculum and operation skill training during the six months. After took the post-test of "RIA Knowledge Mapping and Modelling retrieval system", their result showed a significant improvement.

This also extends to another phenomenon. The users with different computer technical ability; prior knowledge and professional cognition, have different needs in RIA Knowledge Mapping and Modelling retrieval system. In order to make the best of retrieval system, the computer curriculum and operation skill training in advance is essential.

5. Research finding (2): Different users needs and individualized classification

Considering the different users have different prior knowledge, intelligence factor and cognitive style, the researcher designed a Questionnaire of "RIA Knowledge Mapping and Modelling of CM + SS + KOD retrieval system", to test the individuals different needs. The next, using SPSS to analysis, the validity of this questionnaire resulted in a Cronbach's Alpha value is: 0.7074."

Moreover, using ANOVA to analyze the question --- "Do you agree the amplitude of information will cause the user "Cognition overload?" With another variable of: "Do you think the RIA (CM+SS+KOD) retrieval system based on the hypermedia navigation service will enhance the user's cognition and their learning?) --- The Mean Square = 1.343; F = 3.661; Pearson Correlation = 0.264; $P < 0.05$*.

Further more questions are: "Do you think the hypermedia KOD retrieval service, is suitable for personal learning and knowledge management?" --- The Mean Square = 1.852; F= 5.362; Pearson Correlation=0.238; $P < 0.01$**.

79% of all the testees "Strongly Agree" and "Agree": the RIA Knowledge Mapping and Modelling – with CM +Semantic Structure retrieval system, will enhance the user's Web usability & cognition in their learning.

83% of all the testees also "Strongly Agree" and"Agree" that the Modelling with hypermedia KOD service, is suitable for personal learning and knowledge management".

Table 3: Analysis result from ANOVA

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Do you think Concept Map & Semantic structure navigation based on hypermedia service will enhance Web usability in user cognitive in their learning?	Between Groups	4.030	3	1.343	3.661	.015
	Within Groups	41.834	114	.367		
	Total	45.864	117			
Do you think the adaptive and hypermedia on-demand service, is suitable for personal learning and knowledge management?	Between Groups	5.555	3	1.852	5.362	.002
	Within Groups	39.369	114	.345		
	Total	44.924	117			

From the questionnaire test and in-depth interview result, it appears: The three-groups of users have a diverse need with this RIA Knowledge Mapping and Modelling retrieval system, the researcher list the result as below:

(1). How many semantics retrieval sub-links does the interviewee favor?

The expert groups are: 6~10 sub-links; Experienced groups are: 5~10 sub-links; Novice groups are: 6~10 sub-links.

From the testee feedback: Too many semantic sub-links of information (data) in RIA Knowledge Mapping, will defer the time for the user to search and filter the information.

The “Novice groups” also expressed: 5 or 6 web pages (layers) are the most for them.

(2). For three different groups: The different adaptive retrieval preference in RIA Knowledge Mapping and Modelling are:

Expert groups	1. Members of this group with a high level of prior knowledge in CM+SS+KOD, and have a greater ability for comprehension; are able to create the mind-structure by themselves; and there is no need to provide additional Mapping and Modelling navigation support for them. 2. The most suitable way is provide the "Direct guidance" and flexible paths; that will easy for them to find out the specific information, e.g. Search engines and Information on Demand (IOD) retrieval services. 3.Moreover, this expert group will not blindly to use this RIA Knowledge Mapping and Modelling retrieval system, it just depends on the topic or target (if he/she has a clear prior knowledge and certainty over what kinds of topic they want to research), then they may decided to uses the retrieval system or not.
“Experienced” and “Novice” groups”	Typically, this group of members has a low level of prior knowledge in CM+SS+KOD, and has limited comprehension ability about the content; they also lack the domain knowledge and concept structures. The suitable Mapping and Modelling navigation are includes providing the visual cues---with adaptive retrieval structured paths, further to help the novice users integrate their own knowledge, e.g. the “Adaptive sorting of links”; “The adaptation of the guiding map”; and the “Visualized hierarchical maps”.

(3). User Interface design in RIA Knowledge Mapping and Modelling retrieval system:

"Expert" groups	1. Members of this group prefer the “Direct guidance” of “Intuitive observation” design; and regard for the "Integrity”; “Utility” and “Usability” retrieval function. 2. This group of user wish has the control power to predominate and retrieve their information/content. Ex: IOD service.
"Experienced groups"	1. Members of this group prefer the “Simple” and “Explicit” spatial knowledge Mapping design, that including the “Visualized knowledge scaffold” and “Hierarchical structure path”. 2. This group of members do not like the mechanism with numerous and disorderly semantic-sub-links, but should be provided with the “Adaptive presentation” and “Adaptive navigation” retrieval service, which can help them to find the data clearly & organize information quickly.
"Novice groups"	1. Members of this group prefer more”Simple”, “Convenient” and the “Visualize” navigation

	design. 2. Designer should classify the information, and keep the quantity of information moderate. 3. Moreover, provide the "Adaptive navigation" with the structured catalogue and key-word index searching will also helpful. 4. The Mapping and Modelling Interface operation should be more easy and user-friendly for their retrieval. It shouldn't be limited and restrained by the system mechanisms or content provider's personal ideas.

6. Conclusions and Suggestion

RIA Knowledge Mapping and Modelling of "CM + SS + KOD retrieval system" can facilitate the build up of standardization; also it may take advantage of understanding the related or deeper knowledge structure and content. This self-adaptive & object-oriented systems in particular enable the user to easily retrieve the hypermedia metadata and re-use different complex or interesting content.

The researchers' conclusions are: this RIA Knowledge Mapping and Modelling of "CM + SS + KOD retrieval system" has the superior features in that; (1) Extends the characteristic of hypertext; (2) Provides quality of service, adaptive to demand; (3) Overviews the web "shape" of the whole structure; (4) Allows cognitive structure navigation with meaningful learning; (5) Contributes to the cognition memory (based on the UI interactive design and sensorial design); (6) Allows for non-linearity, flexibility and interactive construction in visual communication; (7) Provide the on-demand retrievable & reusable services.

Finally, based on the characteristics of this RIA Knowledge Mapping and Modelling retrieval system, some of the suggestions can be use to help the Web content designer and other researchers in future UI design, such as:
(1) Avoid "amplitude"; "scattered" and "aimless" semantic information design (56% of all the testers "Strongly Agree" &" Agree" too much semantic information, will lead to "Over link" and "Cognitive Overload"; (2) "Direct guidance"-- simple & intuitive is the primary principle; (3) Provide visualizing navigation --- Integrated with multimedia metadata; to promote information retrieval, recording and retention to the related information context; (4) Develop the automated knowledge excavation and organize categorization which provides intelligent / personalized searching and interactive data mining; (5). Provide the classification of User Interface design; (6). Combine the interactive component of the MCU (Multipoint Control Unit) function to extend the retrieval system; (7). Based on the user-centered methods, design personalized channel & personal marketing services, ex: personal VOD & AOD retrieval podcasting. (8). Using the AJAX (Asynchronous JavaScript and XML) and Flash techniques to extend the multimedia video & audio RIA (Rich Internet Application) retrieval service.

7. Reference

[1] Beissner, K., Jonassen, D. H., & Yacci, M. (1993). Using and selecting graphic techniques to acquire structural knowledge. (ERIC Document Reproduction Service N0. ED362151)

[2] Day, G. (1980) "Strategic Market Analysis: Top-down and bottom-up approaches", working paper #80-105, Marketing Science Institute, Cambridge, Mass. 1980.

[3] Duffy, T. M., Lowyck, J., & Jonassen, D. H. (1993). Designing environment for constructive learning. Heidelberg: Springer-Verlag.

[4] Dietrich Albert & Cord Hockemeyer, (1997). Adaptive and Dynamic Hypertext Tutoring Systems Based on Knowledge Space Theory, p.1

[5] La grille d'évaluation de Novak et Gowin (1984), Learning how to learn. Cambridge: Cambridge University Press. pp. 36-37, http://www.uqtr.uquebec.ca/~lamyd/ideateur/evaluation_des_reseaux_de_concepts.htm

[6] Lou Rosenfeld and Peter Morville. (1998). Information Architecture for the World Wide Web. Sebastapol, CA: O'Reilly Media. (ISBN 1565922824)

[7] Nathan Shedroff, (1994) Information Interaction Design: A unified field theory of design, Page 1-2. Retrieved November 2003 from http://www.nathan.com/thoughts/unified/

[8] Pine, J. (1993) "Mass customizing products and services", *Planning Review*, vol 22, July-August, 1993.

[9] The Digital Museum of Nature& Culture. Taiwan. http://digimuse.nmns.edu.tw/index.jsp.

[10] Demonstration website: http://digimuse.nmns.edu.tw

ANALYSIS AND RECORDING OF MULTIMODAL DATA

Mathijs van Vulpen[1], Leon .J.M. Rothkrantz[1, 2], Pascal Wiggers[1] and Alin G. Chiţu[1]

[1]Man-Machine Interaction Group, Department of Mediamatica
Delft University of Technology, Delft, The Netherlands
[2]Netherlands Defense Academy, Nieuwe Diep 8 1781AC,
Den Helder, The Netherlands
{mathijs,l.j.m.rothkrantz,p.wiggers,a.g.chitu}@tudelft.nl

KEYWORDS

Emotion recognition, multimodal, experiments

ABSTRACT

At TUDelft there is a project running on automated assessment of emotions. Prototypes have been developed for extraction of features from facial expressions and speech. To train such systems data is needed. In this paper we report about the recordings of semi-spontaneous emotions. Multimodal emotional reactions are evoked in 21 controlled contexts. The purpose of this database is to make it a benchmark for the current and future emotion recognition studies in order to compare the results from different research groups. Validation of the recorded data is done online. Over 60 users scored the apex images (1.272 ratings), audio clips (201 ratings) and video clips (503 ratings) on the valence and arousal scale. Textual validation is done based on Whissell's Dictionary of Affect in Language. A comparison is made between the scores of all four validation methods and the results showed some clusters for distinct emotions, but also some scatter for certain emotions which depend mainly on the context. Context is not always available.

INTRODUCTION

Emotions are an integral part of our daily rational decision making and communication. To approach the naturalness of face-to-face interaction machines should be able to emulate the way humans communicate with each other. It is the human face that conveys most of the information about our emotions to the outside world. Considerable research in social psychology has shown that besides speech alone, non-verbal communicative cues are essential to synchronize the dialogue, to signal certain emotions and intentions and to let the dialogue run smoother and with less interruptions (Boyle et. al 1994). Non-verbal communication is a process consisting of a range of features including body gesture, posture and touch or paralingual cues, often used together to aid expression.

The combination of these features is often a subconscious choice made by native speakers, and interpreted by the listener. Of all different non-verbal communication means, facial expressions are the most important means for interpersonal communication (Russel and Fernandez-Dols 2004). We learn to recognize faces and facial expressions early in life, long before we learn to communicate verbally.

A human face can supply us with important information. Firstly, it gives us primary information about the identity of the person and its principal characteristics (e.g. sex, race, age). Secondly, the appearance of a human face performs an active role in speech understanding (Benoit and Mahomadi 1994). It is shown that even normal-hearing people use, to some extent, lip reading in order to better understand the speaker. This means that the intelligibility of speech is higher when the speaker's face is visible (Sumby and Pollack 1954). Whenever the amount of sound sources, or the noise, increase the visual information can be very helpful in understanding the message. Thirdly, appropriate facial expressions and/or body gestures also provide additional communicative functions (Graham and Argyle 1975). Often unconsciously, people use nonverbal language (facial expressions, hand gestures, eye gaze etc.) to enrich their dialogue. The other way around, people unconsciously read the nonverbal queues to emphasize what the speaker is communicating. Facial expressions can even be used as a replacement for specific dialogue acts (such as confirmation or spatial specification). The question of how many emotional states we use in our daily communication has yet to be answered. Very little research has been done to locate and recognize emotions other than the six archetypal emotions as described by Ekman (Ekman 2003). Therefore most approaches to automatic facial expression analysis attempt to recognize this small set of archetypal emotional expressions. This practice may follow from the work of Darwin (Darwin 1872) and more recently Ekman (Ekman 2003), who has performed extensive studies on human facial expressions. Ekman found evidence to support universality in facial expressions. These "universal (also referred to as 'archetypal' or 'basic') facial expressions" are those representing the six archetypal emotions: Anger, Disgust, Fear, Happiness, Sadness and Surprise. The recognition of these six archetypal emotions has been done separately

for every modality. The best results are achieved by looking at facial expressions. Further unimodal research has been done for audio, text and gesture recognition. However, all these approaches use mostly their own, self created datasets. This means that comparison between the different recognition and classification algorithms is difficult. As a consequence, there have been very little attempts to combine different modalities and approaches in order to fuse these single modalities into one multimodal emotion recognition system.

The main challenge in multimodal emotion recognition is achieving a high recognition rate in various environments under different circumstances. This means that there can be a lot of noise present in a channel, e.g. occlusion in the face by a hand or glasses, the head is rotated in such a way that it is not perfectly visible or multiple persons are speaking at the same time. The advantage of more modalities present helps in better recognizing the expression based on the fact that these different modalities can enhance each other. A dataset containing all these various environments and different circumstances is yet to be developed. The lack of a widely used multimodal database with data suitable for emotion recognition for unimodal and multimodal systems made us gather data and develop such a database ourselves. The content from this database should be suitable for multimodal emotion recognition systems as well as unimodal emotion recognition systems, and vocal affect recognition. This to make it easier for different research groups to compare their results with other research groups.

All the steps from the creation till the adding of content to the database should be carefully considered in order to get the best basis for all emotion recognition research groups. High speed cameras and sensitive microphones should be used to get the best quality recordings. Reducing the quality of the recordings can be done in a later stage, if necessary. The recording protocol should be suitable for the content to be added and the people being recorded should feel comfortable and relaxed while being recorded. Recording people's response in a foreign language might not be a good approach. The response would be unnatural and the speech part is not fluently spoken, this can cause errors in training speech recognizers. The approach used in (Martin et. al) showed us that non-native speakers give bad recordings as they have to think before responding. Besides this mental issue, the pronunciation of non-native English speakers can cause a difference in pitch variation due to the fact that their native language uses different levels of pitch or intonation. Therefore the database is, for now, filled with native Dutch speakers only, as there are no large numbers of native English speakers available. Furthermore, it is interesting to see whether humans rate images, audio clips and video clips from the same recording as alike. Differences between modalities should be eliminated when rating multimodal information. This hypothesis should be proven by letting people rate the different modalities and then compare the results. Confusion matrices should tell which emotions are alike and which of them are easy recognizable by humans as well as emotion recognition systems.

EMOTIONS RECOGNITION

Emotions are a ritual of our daily routine. They are our main motivators. An emotion gives a human face a very complex structure. Emotions give strength to our way of communicating. Facial expressions are not always a reflection of our emotional state, but can give extra meaning to the content of the message. The human body and specifically the human face provide a lot of conversational information. Facial expressions can supplement text; add an emotional state to the information which helps us to understand a message according to the intention of the speaker. The same holds for vocal affect. Vocal affect can give an emotional load to words or sentences; the intonation or intensity of the spoken utterance gives us clues about the emotional state of the speaker.

The next paragraph presents basic knowledge about emotions, facial communication and emotions that play a role in face-to-face communication. Besides that, some facial gestures can even replace words as e.g. an act of nodding the head can replace a verbal confirmation. In fact, not only the speaker uses facial expressions, but the listener can give nonverbal feedback via facial expressions too. The Paragraph "Facial Expressions Analysis" introduces the Facial Expressions Recognition problem and gives some general approaches to the solution. In this paragraph we also present the Facial Action Coding System (FACS) (Ekman and Friesen). FACS is a facial expressions annotation system which is used in both analysis and synthesis of facial expressions and is used by most researchers who cover this area of research. The process of Vocal Affect Recognition is introduced afterwards. Emotion classification from audio clips has been done since the 1930's. We conclude this section with Multimodal Emotion Recognition. The two fusion methods of the different unimodal channels are explained. Section 3 is about the experimental design and Section 4 reports the results of our experiments.

Emotions

Humans use a daunting number of labels to describe emotion (Cowie and Douglas-Cowie 2001). Therefore most approaches to automatic facial expression recognition attempt to recognize a small set of archetypal emotional facial expressions. This practice may follow from the work of Darwin and more recently Ekman, who has performed extensive studies of human

facial expressions. Ekman found evidence to support universality in facial expressions. These "universal facial expressions" are those representing the six 'archetypal emotions: Anger, Disgust, Fear, Happiness, Sadness and Surprise. Labeling the emotions in discrete categories, such as the six archetypal emotions sometimes is too restricted. One problem with this approach is that the stimuli may contain blended emotions. The choice of words may be too restrictive or culturally dependent too. There is little agreement about a definition of emotion. Many theories of emotion have been proposed. Some of these could not be verified until recently when measurements of some physiological signals become available. In general, emotions are short-term, whereas moods are long-term, and temperaments or personalities are very long-term.

Visualization of emotions

Research from Darwin forward has recognized that emotional states involve dispositions to act in certain ways. The various states that can be expressed are simply rated in terms of the associated activation level. Instead of choosing discrete labels, observers can indicate their impression of each stimulus on several continuous scales. Two common scales are valence and arousal. Valence describes the pleasantness of the stimuli, with positive (or pleasant) on one end, and negative (or unpleasant) on the other. The other dimension is arousal or activation, (Russell 2001). The vertical axis shows activation level (arousal) and the horizontal axis evaluation (valence).

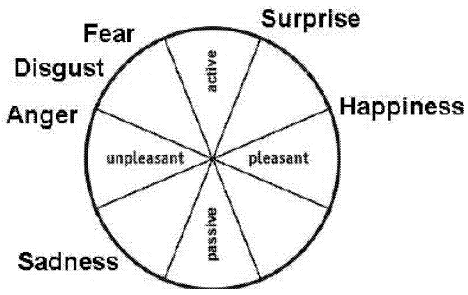


Figure 1: Activation-Evaluation space of six basic emotions.

A circumplex can be viewed as implying circular order, such that variables that fall close together are more related than variables that fall further apart on the circle, with opposite variables being negatively related and variables at right angles being unrelated (orthogonal). We can identify the centre as a natural origin or the neutral state. The neutral state is a condition of readiness to respond. Emotional strength can be measured as the distance from the origin to a given point in the activation-evaluation space. An interesting implication is that strong emotions are more sharply distinct from each other than weaker emotions with the same emotional orientation. A related extension is to think of the six archetypal emotions as cardinal points in that space.

Facial Expressions Analysis

In order to detect and analyze facial expressions we need first to define how to describe these emotional states expressed on our face. The prime example of the codified version of facial expressions is the Facial Action Coding System (FACS), widely used by psychologists. Another way of describing facial activity is to give some quantitative description in terms of geometrical changes of the face. Such geometry changes can be described by using an MPEG-4 standard with its Facial Animation Parameters (FAPs). Other prominent components of the face are the eyes, teeth and tongue. It is the eyes to which people pay a lot of attention during a conversation. The eyeball is generally white with a black pupil positioned in the centre of the visible part of the eyeball, and surrounded by a colorful iris. Every iris is different; it varies in color and structure between individuals. Its task is to regulate the amount of light passing through the lens by controlling the size of the pupil. The teeth and tongue play a minor role in everyday face-to-face communication. The teeth are visible only when the mouth is open, and although they do not attract as much attention as the eyes, they are very important objects in speech processing.

Vocal Affect Recognition

The vocal aspect of a communicative message carries various kinds of information. If we disregard the manner in which the message was spoken and consider the verbal part (e.g., words) only, we might miss the important aspects of the pertinent utterance and we might even completely misunderstand what was the meaning of the message. Nevertheless, in contrast to spoken language processing, which has recently witnessed significant advances, the processing of emotional speech has not been widely explored. To estimate a user's emotion by the speech signal one has to carefully select suited features. Common features include the values of pitch, intensity and spectral features and all of these within a 10 ms or 15 ms time window. The main energy source in speech is vibration of the vocal cords. The vocal cord can vibrate at any given time with any given rate. The rate at which vocal cords vibrate determines the fundamental frequency of the acoustic signal, the so called 'F0'. Traditional as well as most recent studies in emotional contents have used "prosodic" information which includes the pitch, duration, and intensity of the utterance. Variations in voice pitch are considered to have a linguistic function. Pitch features are statistical properties of the pitch contour. The set of spectral features is comprised by

statistical properties of the first 4 formants and the energy below 250 Hz. A formant is a concentration of acoustic energy around a particular frequency in the speech wave which results from the resonant frequencies of any acoustical system. Formants occur roughly at 1000 Hz intervals. Each formant corresponds to a resonance in the vocal tract. The speech intensity depends primarily on the amplitude of vocal cord vibrations which is related to the pressure of the air stream; the larger the expiratory effort, the larger the intensity.

Multimodal Emotion Recognition

The studies in facial expression recognition and vocal affect recognition have been done largely independent of each other. The aforementioned works in facial expression recognition used still photographs or video sequences where the subject expresses only facial expression without an emotionally loaded utterance. Similarly, the works on speech based emotion detection used only the audio information. There are situations where people would speak and exhibit facial expressions at the same time. For example, "he said hello with a smile", here facial expression recognizers may fail due to the fact that the mouth movements may not fit the description of a pure "smile". For computers to be able to recognize emotional expression in practical scenarios, these cases must be handled.

Fusion of Different Modalities

Multimodal fusion is the integration of the information present in individual input modalities, here audio and video features carrying information about emotions. Integrating modalities requires understanding how people use their various senses to perceive and interact with the world around them. It will depend on knowledge of the natural integration patterns that typify people's combined use of different input modes. This means that the successful design of multimodal systems will require guidance from cognitive science on the coordinated human perception. However, exactly how the fusion of audio and visual information takes place in human perception is not yet answered. In particular, the existing studies do not agree on the central question of at which stage the fusion occurs.

The various existing models can be roughly categorized into two groups:

- Early integration models.
- Late integration models.

In the perspective of the early integration models, or feature-level fusing, fusion takes place before the recognition stage. These methods typically use only one classifier, combining the features from the different modalities as its input. The other class of methods carries out fusion in the decision-level, which are referred as late integration or decision fusion methods. In these methods the likelihood scores of the single-modality classifiers of each modality are calculated independently from one another, and the fusion is carried out by combining the decisions given by the parallel channels.

THE EXPERIMENTAL DESIGN

The experimental condition

The experiment is done in a closed room with good lighting conditions. A good lighting condition means that there is enough diffuse light to leave no shadows on the participants face. The camera is focused on the participant and the height of the camera is just right for lip reading. A consequence of this is that the height of the camera must change for every participant in order to get a perfect frontal view, but this is easily solved by placing the participant higher or lower in the chair. The background behind the participant is covered with a dark color, preferably blue or green.

In the case of video data recording there are a larger number of important factors that control the success of the resulted data corpus. Hence, not only the environment, but also the equipment used for recording and other settings is actively influencing the final result. The environment where the recordings are made is very important since it can determine the illumination of the scene, and the background of the speakers. We use a mono-chrome background so that by using a "chroma keying" technique the speaker can be placed in different locations, inducing in this way some degree of visual noise.

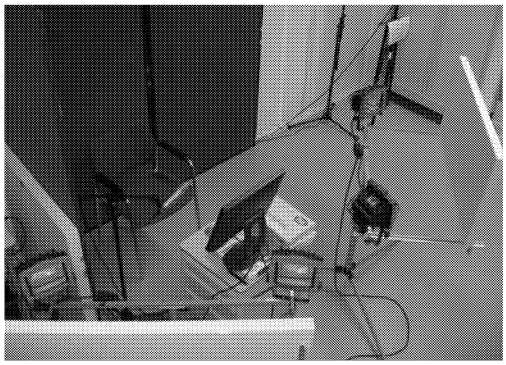


Figure 2: Overview of the room setting.

The Procedure of the Experiment

The database should ideally contain only genuine expressions of emotions. However, as the database should also consist of high-quality video samples (with constant illumination, background, head pose, etc…) to be useful for practical applications, the choice that was made was to get as close as possible to spontaneous

emotions, while keeping at the same time a fully controlled recording environment. To achieve this goal it was chosen to use pre-defined answers for each situation. After a consent form was signed, the experiment could start. The average time the experiment lasted was 45 minutes. For each emotion the participant was asked to listen carefully to a short story and to 'immerge' themselves into the situation. Once ready, the participant may read, memorize, and pronounce (one at the time) the five proposed utterances, which results in five different reactions to the given situation. The participants are asked to put in as much expressiveness as possible, producing a message that contains only the emotion to be elicited. To obtain the participant's facial expressions the experimenter should instruct the participant to perform the pure facial expressions. The procedure of the experiment is as follows:

1. The participant is told about the short stories to which he or she has to react to. The corresponding emotions are explained beforehand, so the participant knows which emotion to express.

2. The participant gets a list with responses corresponding to the expected expressed emotion.

3. The participant is asked to sit in front of the frontal camera, while the profile camera is positioned to the left of the participant. The distance between the cameras and the participant is about two meter.

4. The participant listens to or reads the short story and is asked to imagine being in this situation.

5 The experimenter gives the order to capture the emotional expressions of the participant.

6 The participant is then asked to react with each of the five pre-defined sentences.

7 If the participant's performance is not ideal, repeat step 4 to 6. The repetition times are under control.

8 After the experiment is done, thank the participant for the cooperation.

Repeat step 4 to 7 for the next designated emotional expression, until the participant has displayed all of the emotional expressions.

With this procedure we let our participants take multiple sessions. This way we can fill the database quickly with many different recordings. The goal is to create a balanced database with respect to gender and age. A constant adding of recordings to the database will ensure the expansion of the data corpus and will add more diversity to the recordings.

VALIDATION

All recorded data has to be validated and there are several methods to do so. In the next sections, we will explain how we offered the data for validation and how we processed the results. We split up the validation into three different parts, this to see if there are differences in the validation of the various modalities. We offered from the recordings, the image showing the apex of the emotion, we offered only the audio clip from the recording and we offered the whole recording as a video clip. We made a website where users could validate the data on two dimensions, valence and arousal. In more detail we requested the users to score a sample image, sound clip or video clip on these two scales. Note that also non-native Dutch speakers could, and did, validate the images. We therefore made the language of this web site English. How many non-native Dutch users there were is not known. We ensured the users privacy and did not record any personal data, except their IP address. This in order to know how many distinct users participated in our validation process. Most users are naïve users and therefore they had the opportunity to look at an example validation before starting their own validation process. Examples of the validation processes of images, audio clips and video clips are visualized in Figure 3.

There are several methods to validate and annotate the recorded clips. Experts can annotate the clips to their best knowledge of the domain, but naïve users can give labels to the recordings too. As long as there are enough different people who annotated the clips, the 'consensus' of the broad public can be extracted.

Image based validation

In order to get a lot of ratings in a short time the validation of the images was done on-line. In order to quickly start validating the recorded clips, images showing the apex of the expressed emotion from the clips we extracted. These images were then displayed at a special developed website for validation. Users could click on 'validate Images' to start validating images right away. There was no limited amount of images shown and the displayed images were selected in a random order from all available images in the database. These images are all extracted from the recordings of one person. In a matter of days more than 30 people had been rating images and the number of ratings exceeded 1200. The blended emotion, like Anger-Surprise, was one for which we could not extract a single apex image from the clip. This is because the clip clearly showed two emotions being blended very fast after one another. The surprise expression of the emotion came after the anger expression. A list containing all of these emotions can be found in Section 5, note that the image for Anger-Surprise was not validated and that there was an image present expressing neutral face.

Audio based validation

Validation of the various audio clips was done on the same website too. Here the users first listened to an audio clip taken from a recording and then they had to rate what they thought about the expressed emotion on the 9-point scales. We presented a number of 105 different audio clips, divided into 21 emotions and per

emotion 5 different audio clips. These audio clips were all taken from one participant. The screen presented to the user differs not much from the screen presented for the validation of images, this in order not to confuse the user. If a user can validate and rate an image, they know how to do this for an audio clip too. The user is in control of the playback of the audio clip. If necessary the user can play the audio clip again and again. When the page was loaded the audio clip automatically started to play.

Video based validation

For the video based validation the same principal was used as for the aforementioned image and audio validation. Here the users could control the playback of the video. When the video was loaded the user could start playing it at any time. The videos were constructed from the images recorded with the high speed camera to a format presentable for average online access speeds. We then converted the movie to a flash movie. These so called .swf files are suitable for online presentation and commonly used in websites like youtube.com. An overview of the showed video clip and its controls is shown in Figure 3 below.

Text based validation

For the text based validation we used the Whissell database to look up the corresponding scores for the 21 different displayed emotions. We selected 21 emotions, with corresponding labels. We assume that the answers retrieved from the database show corresponding outcomes with the ones retrieved from the validation process. The whole database consists of more than 10.000 words and every word is scored by many respondents. We looked up the scores of our selected 21 emotional labels and where necessary substituted unknown words to the database with similar ones. The Whissell scores are based on a 3- point scale so these scores were converted to the 9-point scale used in the previous explained validation processes. The text based validation was not done by the developed website. A second web interface was created to extract the values from the database. An example, here "Interest", is shown in Figure 3.

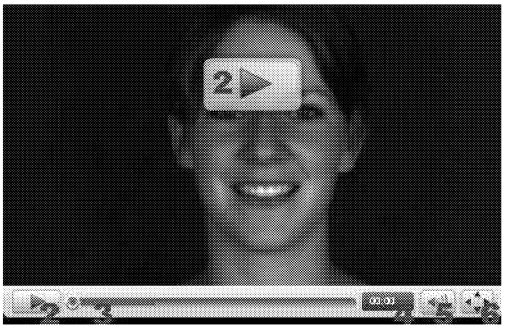


Figure 3: Explanation of flash content with controls.

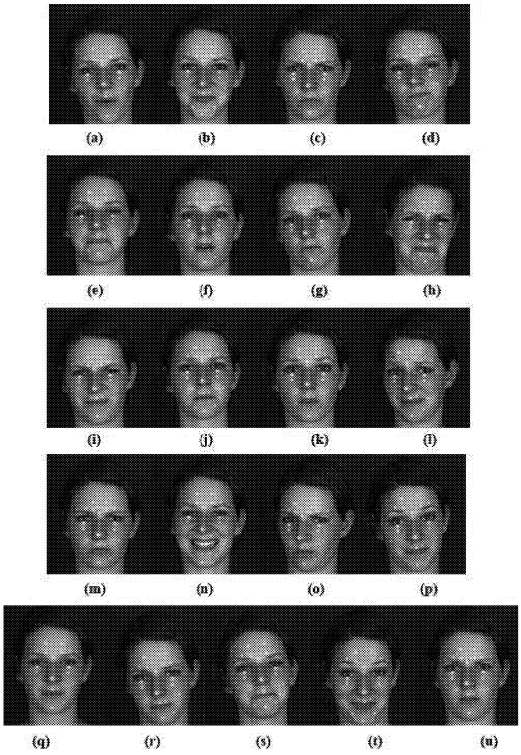


Figure 4: Display of all 21 emotional facial expressions.

RESULTS

The used validation approach was to annotate the images/clips with a value for valence (positive vs. negative) and a value for arousal (passive vs. active).

These validation results could then be drawn on the two axes. We can look at the validation result in several ways. We can look at the different modalities separately. We can look at the average of all modalities combined in the video. If we look at the modalities separately we expect users to score the displayed emotion incorrectly, due to the lack of extra information we usually get via other modalities. However, if we look at the validation of the videos, where all modalities are present, we expect the results to be similar. More information is present, such that users can give a more precise score for the emotion. A more detailed description of validation results per modality is given in the next four paragraphs. Validation tables are given for the validation scores per modality separately too. A total of 62 unique users participated in the validation of the different media displayed on the web page. As mentioned before, not every user that validated images validated audio and video. This is because the language

of the spoken text is Dutch and we also asked foreign users to validate the images.

Image based validation

A total of 49 unique users validated the images present on the website (Figure 3). It proves that the analysts were able to clearly distinguish the 21 that different facial expressions (Figure 4). If the users score these expressions widely spread across the valence and arousal axis the prototype should be able to correctly classify these 21 expressions. A distance matrix was computed, where the distances between the facial feature vectors of both persons are compared. However, the results from the image validation were not as expected. The expressed emotions were not over expressive, so users hardly rated the expression shown in the images as very aroused or activated. The expression of happiness has a very good (or what was expected) validation, however the validation of contempt is way off and is even validated as a positive expression. A comparison between the different validation methods is visualized in Figure 5. An overview is given in Table 1, below.

Table 1: Validation results for image validation.

		Valence	Arousal			Valence	Arousal
1	Admiration	0.81	0.19	*12*	Fascination	0.03	-0.74
2	Amusing	1.92	0.58	*13*	Fear	-1.17	0.23
3	Anger	-2.16	0.39	*14*	Furious	-2	-0.06
4	Anger (Surprise)	Not validated		*15*	Happiness	2.99	2.12
5	Boredom	-0.68	-1.02	*16*	Indignation	-1.47	-0.2
6	Contempt	0.7	-0.65	*17*	Interest	1.38	0.2
7	Desire	0	-0.69	*18*	Neutral	-0.12	1.4
8	Disappointment	-1.42	-1.1	*19*	Sadness	-0.69	-0.98
9	Disgust	-2.83	1.64	*20*	Satisfaction	0.79	-0.34
10	Dislike	-1.61	0.44	*21*	Surprise (Pleasant)	2.48	1.52
11	Dissatisfaction	-0.77	-1.25	*22*	Surprise (Unpleasant)	-1.54	0.09

Auditory validation result

A total of 9 unique users validated the audio clips present on the website, till now. The presented audio clips were the audio tracks from the video clips present at the website too. A comparison between the different validation methods is visualized in Figure 5. An overview is given in Table 2, below.

Table 2: Validation results for auditory validation.

		Valence	Arousal			Valence	Arousal
1	Admiration	2.5	1	*12*	Fascination	0.75	1.25
2	Amusing	1.43	0.57	*13*	Fear	-1.86	2.57
3	Anger	-2.91	2.82	*14*	Furious	-3.29	3.14
4	Anger (Surprise)	-1	2.33	*15*	Happiness	2.5	2.25
5	Boredom	-1	-1.67	*16*	Indignation	-0.33	-1.17
6	Contempt	-0.4	-0.6	*17*	Interest	2	1.75
7	Desire	2.25	2.5	*18*	Neutral	Not validated	
8	Disappointment	-2	-0.83	*19*	Sadness	-2.71	-0.43
9	Disgust	-1.75	2	*20*	Satisfaction	3	-3
10	Dislike	-2.17	0.17	*21*	Surprise (Pleasant)	3	3
11	Dissatisfaction	-1	0	*22*	Surprise (Unpleasant)	0.67	1.33

Multimodal validation results

A total of 14 unique users validated the video clips present on the website, till now. Remarkable is the fact that these users gave the emotion 'fear' a negative arousal and all the other validation method scored a positive arousal for fear. A comparison between the different validation methods is visualized in Figure 5. An overview is given in Table 3 below.

Table 3: Validation results for multimodal validation.

		Valence	Arousal			Valence	Arousal
1	Admiration	1.5	0.43	*12*	Fascination	1.24	0.33
2	Amusing	3.65	3.12	*13*	Fear	-3.09	-1.64
3	Anger	-3.41	2.53	*14*	Furious	-2.77	2.46
4	Anger (Surprise)	-2.86	1.07	*15*	Happiness	3.83	3.33
5	Boredom	-1.31	-2.13	*16*	Indignation	-0.45	-0.27
6	Contempt	0.6	1.4	*17*	Interest	2.47	1.42
7	Desire	2.61	1.56	*18*	Neutral	Not validated	
8	Disappointment	-1.93	-1.67	*19*	Sadness	-2.43	-1.79
9	Disgust	-3.55	0.09	*20*	Satisfaction	2.57	1.93
10	Dislike	-2.44	0.38	*21*	Surprise (Pleasant)	2.92	2.25
11	Dissatisfaction	-1.35	-0.41	*22*	Surprise (Unpleasant)	-3.29	1.07

Textual validation results

A total of over 200 unique users validated the words in the Whissell database. The Whissell database consists of over 10.000 words for which these users gave a score for the valence, arousal and imaginary scale.

In our research we do not use this imaginary scale. The Whissell database gave the following coordinates for the evaluation and arousal space, showed in Table 4 below. Note that the Whissell scale is a 3-point scale, rating from 1 to 3, where 1 is the lowest value and 3 the highest. To compare these Whissell values we have transformed these outcomes to the 9-point scale we used. The numbers in front of the emotions in the table correspond to the red number shown in Figure 5.

Table 4: Whissell's scores for the emotional words used given on a 9-point scoring scale.

		Valence	Arousal			Valence	Arousal
1	Admiration	2.00	-0.89	*12*	Fascination(Fascinating)	3.20	2.00
2	Amusement(Amusing)	4.00	2.50	*13*	Fear	-4.00	1.45
3	Anger	-4.00	3.56	*14*	Furious(Fury)	-2.40	2.29
4	Anger (Surprise)	Not validated		*15*	Happiness	4.00	3.20
5	Boredom	-4.00	-2.55	*16*	Indignation(Resentment)	-2.67	-0.44
6	Contempt	-1.71	-1.00	*17*	Interest	2.67	-3.00
7	Desire	0.57	1.50	*18*	Neutral	Not validated	
8	Disappointment	-1.33	-0.57	*19*	Sadness(Sad)	-2.50	-2.29
9	Disgust	-2.5	-0.50	*20*	Satisfaction	2.22	-3.00
10	Dislike	-3.43	-1.71	*21*	Pleasant Surprise	2.91	1.00
11	Dissatisfaction(Disapproval)	-2.55	-0.67	*22*	Unpleasant Surprise[23]	-2.91	1.00

Comparison of validation results

If we compare the results acquired with the different validation methods used and described in the previous paragraphs we can draw intermediate conclusions. A simple overview can be generated if we combine all four previous tables into one figure. Each validation method is represented with its own color; this is the same color as the numbering in the previous four tables.

The resulted figure, Figure 5, is shown below. In the ideal case we should see clusters of the same numbers close to each other. But this is not the case for many of the 21 emotions. For example, the number 6, representing the emotion 'contempt' is scattered over three quadrants of the axes. The image validation gives this emotion a positive valence and a negative arousal, the auditory validation a negative valence and a negative arousal, the video validation a positive valence

and a positive arousal, and the textual validation give contempt a negative valence and a negative arousal.

This example shows that a lack of information, due to the separation of the unimodal channels can cause the user to misplace the intended emotion. A positive example is, for example, the number 8. This number represents the emotion 'disappointment'. It can be clearly seen that all four number eights lay close together. This means that the validation of this emotion does not depend much on multimodal information. Other clear clusters are the numbers 15, 21, 3 and 14, representing happiness, pleasant surprise, anger and furious accordingly. These very active emotions were validated very closely by all validation methods, showing that the distinction between active positive and negative emotions is not as subtle as with other emotions.

Table 5: Overview of position of emotions per quadrant (RED: Image validation, GREEN: Audio validation, BLUE: Video validation and PURPLE: Text validation.)

		1st Quadrant	2nd Quadrant	3rd Quadrant	4th Quadrant
1	Admiration	IAV			T
2	Amusement	IAVT			
3	Anger		IAV		T
4	Anger (Surprise)		AV		
5	Boredom			IAVT	
6	Contempt	V		AT	I
7	Desire	AVT			I
8	Disappointment			IAVT	
9	Disgust		IAV	T	
10	Dislike		IAV	T	
11	Dissatisfaction			IAVT	
12	Fascination	IAVT			
13	Fear		IA	VT	
14	Furious		IAVT		
15	Happiness	IAVT			
16	Indignation			IAVT	
17	Interest	IAV			T
18	Neutral		I		
19	Sadness			IAVT	
20	Satisfaction	V			IAT
21	Pleasant Surprise	IAVT			
22	Unpleasant Surprise	A	IVT		

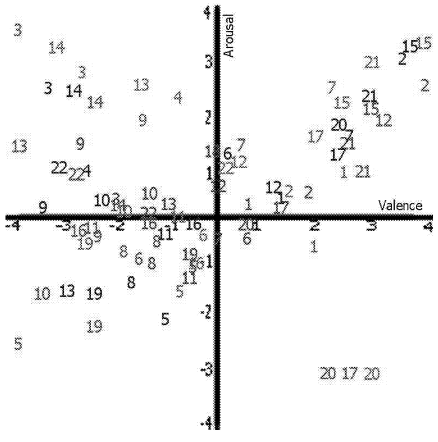


Figure 5: Comparison of different validation methods. RED: Image validation, GREEN: Audio validation BLUE: Video validation and PURPLE: Text validation.

REFERENCES

[1] E. Boyle, A. H. Anderson, and A. Newlands, "The effects of visibility on dialogue and performance in a co-operative problem solving task," Language and Speech, vol. 37, pp. 1- 20, 1994.

[2] J. A. Russel and J. M. Fernandez-Dols, "The psychology of Facial Expression," vol. 9, no. 3, pp. 185-211, 1990.

[3] K. S. Benoit and C. Mohamadi, "Audio-visual intelligibility of French speech in noise," Journal of Speech and Hearing Research, vol. 37, pp. 1195-1203, 1994.

[4] W. H. Sumby and I. Pollack, "Visual contribution to speech intelligibility in noise," Journal of the Acoustical Society of America, vol. 26, pp. 212-215, 1954.

[5] J. A. Graham and M. Argyle, "A cross-cultural study of the communication of extra-verbal meaning by gestures," International Journal of Psychology, vol. 10, pp. 67-67, 1975.

[6] P. Ekman, "Emotions Revealed". New York: Times Books, 2003.

[7] C. Darwin, The Expression of the Emotions in Man and Animals. 1872.

[8] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The eNTERFACE'05 Audio-Visual Emotion Database," in , Atlanta, Apr. 2006.

[9] P. Ekman and W. V. Friesen, The facial action coding system: A technique for measurement of facial movement. 1978.

[10] R. Cowie and E. Douglas-Cowie, "Emotion Recognition in human-Computer-Interaction, "IEEE Signal Processing Magazine, Jan. 2001.

[11] J. A. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol. 39, pp. 1167-1178, 1980., Vancouver, Canada, 2001.

MULTIMEDIA STORYTELLING

THE MAKING OF AN INTERACTIVE DIGITAL NARRATIVE – *INSTORY*

Helena Barbas
Nuno Correia

CENTRIA and DEP/FCSH
CITI and DI/FCT
Universidade Nova de Lisboa, Portugal,
E– mail: {h.barbas|nmc}@fcsh.unl.pt

KEYWORDS

Multimedia storytelling; Interactive fiction; mobile devices; narratology; cultural heritage; serious games;

ABSTRACT

This paper describes the problems that had to be faced during the elaboration of an interactive narrative for the ***Instory*** project (http://img.di.fct.unl.pt/InStory/) directed by Prof. Nuno Correia. The project had the goal of defining and implementing a platform for mobile and cinematic storytelling, information access, and gaming activities, in Quinta da Regaleira (World Heritage) in Sintra, Portugal. The system is driven and validated by a set of fictional threads that are centred on the exploration of physical spaces (the real world, in real time). The development of a narrative was naturally constrained by the environment which raised some practical and theoretical issues in what regards the literary strategies involved. ***InStory*** received the PMA Award (2006) for best Portuguese multimedia project web– mobile.

INTRODUCTION

The ***InStory*** project had the goal of defining and implementing a platform for mobile and cinematic storytelling, information access, and gaming activities (Correia et al. 2005a), initially using a PDA and later mobile phones, in Quinta da Regaleira (World Heritage) in Sintra, Portugal.

Technically, the platform has a quite supple computational architecture that integrates heterogeneous devices, different media formats and computation support for different narrative modes and gaming activities (Martins et al. 2005).

However, the making of a first interactive narrative for the above project, in the terrain – where the fictional and real universes blend – raised a series of theoretical, literary and narratological issues, shaking the concepts and terminology that have been the basis of digital fiction studies.

The physical space in Regaleira is very rich and complex. It was built in the historical centre of the village of Sintra in the early nineteenth century. It materializes the mythical and magical dreams of its deceased owner, António Augusto Carvalho Monteiro (1848– 1920), depicted by the art of his friend the Italian architect and theatre scenographer Luigi Manini (1848– 1936). The space has museological characteristics, embodying the so called Portuguese Romanticism, having become a symbolic tribute to some famous epochs and figures (the Templar's, Dante, our epic Camões, etc.). It has a palace– museum, a chapel, and gardens full of artistic and mythical elements – wells, caves, lakes, towers, sculptures, paintings, tiles, etc. This vast scenario had to be used as a necessary map in which the events of the story would have to take place.

Figure 1. Quinta da Regaleira, Sintra

The project explores the social aspect of shared narratives and activities, having in mind that technology can provide new innovative approaches to social participation in different types of events, being it artistic or cultural (Correia et al 2005b).

The idea of an interactive format for mobile storytelling was to guide the user through her visit to Regaleira. To help her map out the vast and intricate geographical area and show the thematic or historical places that could interest her most amongst the nearly 20 sites available.

Out of the geographical map emerged a web of possible virtual paths. Within this maze was selected a number of specific points/nodes where some of the paths met or intersected. In these nodes the user is allowed to change course and, at the same time, vary the type of game or narrative. For that, a multiplicity of routes has been anticipated allowing the potential multiplication of different stories. So, a basic structure was created, one that could be reproduced in several nodes, as the embryo of a future and more complex fractal structure (Holtzman 1997).

This web of possible paths could become the basic framework for the creation of a future role-playing interactive game, in a virtual environment (reproducing the existing Regaleira space), equally inhabited by avatars and/or humans. A game with a hunting narrative was developed to start with.

This practical experience with the ***Instory*** project has helped to clarify the way in which some of the terms pertaining to narratology can be employed – in particular the distinctions between «story» and «discourse» (Barbas and Correia 2006). This double model is presently being studied by other researchers from diverse perspectives (Young 2006). They are the trick of the trade used by writers to catch the reader's interest, and become crucial to the creation of suspense; they are also the signature of the storyteller, the gap through which creativity erupts. Concerning ***InStory***, creativity lies, partially, in the route taken by the user, as an outcome of the several nodes the she chooses to cross. Partly because there is a previous set of story threads necessary to build the system, and validate it. In each spot of the terrain, the performance of the narrative sequences and general structure was tested by the members of the project.

The previous plots were centred on the exploration of physical spaces – the real world, in real time. And their development was naturally constrained by that very same environment. This interfered with other literary tools to promote the immersion of the reader– user – analepsis / flashback and prolepsis / flash– forward. In the terrain they do not exist, even if we consider the all geographical space as a huge narrative map, a kind of gigantic prolepsis in the beginning of the story.

IF AND TRADITIONAL NARRATIVE MODES

We are in the middle of the digital revolution. Computers have been being used to create stories for a very short period of time yet. Writers are experimenting with this new instrument, wobbly if we consider the pace at which computers and software change, resurfacing every two months with fresh possibilities. From the inert book the creative act transmigrated to multimodality (Nigay and Coutaz 1993), to hypermedia, a plural and very particular communication channel. The emerging mobile technologies took storytelling onto a whole new level (Correia et al 2005b).

Regarding the narrative practices, we are truly experiencing something quite new, and are not yet fully aware of all the potential it has to offer. Studies are being carried out concerning the so called Narrative Intelligence with some promising results (Mateas and Senghers 1999). Yet, the brain that humans use to fabricate stories, the strategies used to organize and transmit them, have not changed that much (Pereira 2008).

Theoretically, the new ways for telling stories are being evaluated and limited by the old traditional modes – the remains of poetics for genres, i.e. – and the most recent of the methodologies, the one that seems the most able to encompass the new needs, is narratology (Monfort 2007; Douglass 2007).

The quick development and increasing use of new media technologies demand a revision of the literary taxonomy, as well as a reconsideration of the instruments to evaluate the new productions.

Interactive Fiction (IF) – narrative or drama – has been substantially discussed as a new form of art related to AI-based experiences and narrative intelligence (Mateas and Senghers 1999). In practice, there has been a considerable technical progress in building elaborate plots and quite believable fictional characters.

However, on of the main issues about the IF theoretical framework has to do with satisfactory terminology. Cyber-artists and critics use a vocabulary borrowed from the realm of literature and film practices (plot, character, perspective, narrator), and from game author's specific language (character player, drama manager). In general it is used in an inaccurate way, without considering all the variants that enriched these concepts, such as narratological studies. From these mishandling results a multiplicity of terms and some misunderstandings in what concerns the theoretical speculations and approaches to digital fiction.

The practical experience with the ***InStory*** project has helped to understand and clarify the way in which some of those literary notions can be employed. It is an extreme situation in what regards the making of a mobile ID story, the exploitation of the usual narrative strategies, and the user's agency and role, using a mobile device.

Being a mobile cinematic and digital project, ***InStory*** is a debtor to its predecessor documentaries (Davenport 2003; 2005). The user will be able to interact with the central server, sending, receiving and asking for any kind of media elements – video, images, games, music, or messages. These elements are directly related to the environment and, in some cases, can describe scenes occurring in that same exact locale (Pan 2004) becoming a kind of posterior journalistic account. The main difference between this kind of mobile storytelling and the traditional narrative modes (oral, written, cinema, and theatre, hypertext, or even ergodic literature) results from the blending of the fictional with the real universes, and in the fact that the narrative is not utterly based on a previous script.

The user is actually experiencing the story immersed in the real world, in real time, and she has the possibility to decide which way to go. Ideally, she can choose her own actions, and her virtual characters' behaviour, select her pathway towards the outcome.

The development of the basic IF content was naturally ordained by that very same environment. So, the main restrictions, narrative wise, were physical and in direct opposition to each other – the vastness of the geographical space and the amount of information available, versus the boundaries imposed by the mobile apparatus features and screens (640x480 pixels for a PDA). Experiences made with this kind of narratives – *MIT in Pocket*, or *15 Minutes*, i.e. (Pan 2004) – give the user a task to be accomplished within a predetermined period of time. Quinta da Regaleira' terrain is very irregular, and the visibility between spots is very poor, so time could not be the motivation issue to make the user go from one place to another.

THE MAKING OF AN INTERACTIVE STORY

The ambiguity of the concept narrative has been widely discussed (Barbas and Correia 2006), having in mind the French structuralists (Genette 1996) and Linguists (Benveniste 1974), the first to propose an opposition between «story» (content plane) and «discourse» (expression plane). Both concepts were also explored by Russian formalists (Todorov 1983) and inherited by American scholars (Prince 1988). These formalistic narratological notions, normally employed as analytical tools, proved to be very useful for operational purposes.

The distinction between «story» and «discourse» could be projected over the geographical map of Regaleira in the design of the possible routes to be taken by the ***InStory*** user, but with some particularities.

The events («story») were allocated to some specific spots, from which the narrative was built, and had to respect the architectural and cultural motives of the geographical space. But the order of the events (the «discourse») is determined by the material route the user wishes to take. That is, the «discourse» is mainly created by the user in the moment she chooses which way to go – her creativity lies in the route taken, results from the several nodes crossed. This route also determines the length of the story, the number and variety of the episodes experienced and the locale where it ends.

Several objectives were targeted: to integrate the possible discontinuity of the fictional sequences and, at the same time, to maintain the illusion of narrative continuity; to advance the user's agency, truly allowing her to play a part in the story development and its conclusion; to have in mind that the user could abandon the story at any moment; to consider the possibility that the user might return to a spot that already been visited.

The user's progress is instigated by a number of approaches: a direct instruction via message (text, audio or image); tests, games or competitions – allowing her to accumulate or lose objects, to continue or change her course, to go up one level in the story/game and gain points; to partake a quiz with multiple choice questions. Also, to solve puzzles/enigmas that she has to decode retrieving information from the geographical spot occupied at each moment.

Theme

Having in mind all the above issues, and that ***InStory*** is also a mobile cinema project, it was decided to use the most obvious elements at hand in the physical space.

The close examination of the surroundings, of the architectural places (turrets, benches, wells, lakes and grottoes) together with the motifs available in the decors, suggested hunting as a first possible organizing theme. The deceased owner of Quinta da Regaleira was called Carvalho (Oak) Monteiro (Huntsman). And he had inscribed the surroundings with all the variants from the elements of his name – oak leaves, hunting scenes and wild animals. There exists a «hunting room» in the palace with birds, boars and stags; there are two huge tile panels with renaissance hunting scenes adorning the walls in the main entrance gate; a major stained glass window in the chapel depicts an old Portuguese legend about a horseman – Dom Fuas Roupinho – chasing a stag.

The relevance of the hunt theme opened an easy and promising way. It could include all the potential routes. It provided a modest beginning and allowed future expansions into higher levels of complexity. It could even include posterior hunting motives: treasure hunt, ghost stories, or even a detective story.

Following the tale morphology theory (Propp 1969), the first incentive for action can be a need to be satisfied. Here, «hunger» was chosen to be the initial «lack of something» to be fulfilled during the quest of the user-hero.

Characters vs. players

Literary studies on character have become even more complex with the interference of Information Technology. There was already an inconsistency between the general concern with exteriority (classic rhetoric) and absolute interiority (romantic tradition) not unravelled by structuralist studies (Barbas 2006). However, some vital assertions were made: the difference between human being and character is absolute; character is an «open concept».

The problem of «consciousness» and «emotions» concerning digital characters has already been sketched (Barbas 2006), being these evaluated the same way they were regarding theatre or cinema actors. Also, the problem of characters verisimilitude implies the use of social rules and cultural norms specific to each time and space they have to inhabit (Si 2006).

The studies about agents and avatars in IF have not yet properly considered these problems, mainly because they bring into play ready-made characters as an example, or create them for each specific purpose. Interactive stories, M.U.D.´s, and role-play games (*Façade*, *Sims*) use general software to develop narrative situations as a whole, or provide a catalogue of ready-made images to choose from (*Second Life*). Also, the relationship between character and plot is not symbiotic as has been proposed (Barbas and Correia 2006). In fact, plot needs agents for its action, but agents are interchangeable inside a plot.

For the ***InStory*** narrative there were created several virtual characters, also inspired by the figures represented in the décor. They are very simple, but can become more complex if coupled with AI (Young 2000) decision models (Dell'Acqua et al. 2006) and preference revision (Pereira et al. 2009).

They can function autonomously as narrators, or as avatars of the user. The heroes are a Hunter (*Jorge*) and an Amazon (*Diana*). The White Lady (*Dama Branca*), who is an Eco-Vegan animal protecting figure, and the dog, the Greyhound (*Galgo*), play a double role of adjuvants and/or opponents to the hunter(s). The Wild Boar (*Javali*) is the victim, representing the other entire kill (stag, lion, and birds).

In ***InStory*** the heroes (*Jorge/Diana*) are the main player-characters (PC); the user can choose her avatar's gender in the beginning of the game. Their points-of-view are, respectively, limited to their functions and actions.

Narrator vs. meta– author

In traditional narrative the narrator is any of the entities responsible for telling the story, and the author is the human having created it. For instance (Todorov 1983) the author has multiple duties and functions, being directly responsible for characters, and the way the reader receives the information he wants to give him. In IF the author is more a «facilitator» (Lambert 2002), an agent with more knowledge than the other characters, sometimes defined as «drama manager», replicating some of the tasks of the theatre/film directors.

InStory is an authoring environment filling the existing gap between the creator and the system. The authoring process (Correia et al.2005a) is simplified through a dragging and dropping interface that includes the story/game components. This prevents the need for the user to have previous programming/advanced computer skills. But it has also pre-defined story content.

In IF the author may manipulate or encourage the user into making particular choices in order to advance the plot; but the user is not the author as the number of options is naturally limited. In ***InStory*** the «literary author» is a member of a crew; he has a role similar to a scriptwriter for film; he shares the authorship with the programmer and all other technicians that cooperate in the making of the story, but also with the user.

In open and pluri-dimensional IF systems the user is physically and mentally active within the realm of the work, she has a practical role in triggering it. This integrative process leads to question narrative paradigms. It attests the unsuitability of traditional terminology to define new practices, and demands the reformulation of literary concepts and relationships.

Within the ***InStory*** system the author and narrator functions are fragmented and/or shared. The PCs can be interposed by the White Lady, who may suggest alternative behaviours; they share the role of the reader/audience, as is their task to «interpret» the tale they are making; and consequently, they also partake in authorship, as they partially control the main plot. Besides, the system allows users to contribute with new data – text or images – that can be contextually integrated into the story threads. Also, there is an «unusual narrative voice» in the character of the dog that has the ability to lie to the heroes.

The plot

The heroes – *Jorge/Diana* – are hungry. They have to go hunting in order to get something to eat. They need to find a weapon and some ammunition in the surroundings, and decide which animal they want to chase (wild boar, stag, or birds). The hunting starts with a Quiz. If the user answers correctly the first set of questions, she will get three bullets. She can also look for floral elements (acorns, i.e.) to swap for apples. She can ask the Greyhound for help, but the dog might lie.

The Dog is friends with the hunter(s), and with the wild boar. It will have to decide which one to help. If the hunter kills, the dog will have some boar meat. If the wild boar escapes, it can have some apples as reward, and exchange them for bones.

The White Lady will try to stop the hunting, either offering fruit (apples) in exchange for bullets, or helping the boar to hide/escape. In the event of her losing, she will have to eat a wild boar pie in the end.

The Boar knows it is going to be chased, and has to escape and hide.

In the end there is a Virtual Banquet of apple pies, or boar stew, where points and or ammunition are exchanged for food.

For the moment, the choice of actions belongs to the user, but this generic plot can also be transformed by future use of AI genetic algorithms, or explored if converted to a virtual game. This would decrease the decision power of the user, reducing her authoring capacities, but this loss would be compensated with the element of surprise when in the terrain – suspense – which is one of the properties of linear/traditional storytelling that the audience misses the most.

Structure

Having in mind also the structuralist and post-structuralist narrative studies (Barthes 1996; Greimas 1996), the idea of a basic structural narrative sequence that could be reproduced in each different context was adopted. The aesthetical proposals for virtual space (Holtzman 1997) inspired the elaboration of a fractal model that, being transformed by each environment, could be multiplied in secondary levels.

Accordingly (Greimas 1996), each of the narrative sequences is composed by three «functions»:

F.1 – beginning of an action – the hunt starts;

F.2 – execution/non execution of the action (bifurcation) – where the user can ask for help from another character, or improve her chances of winning by answering a Quiz.;

F.3 – end (success/failure) of the action – advancement to another sequence, or level of the game.

This basic sequence is replicated, yet keeps changing because of the surroundings. The main difference towards traditional narrative modalities is that each sequence has to complete itself – open and end – at the same place. This means that the climax of each event, the raising and falling actions, have to occur almost immediately. As a consequence, it becomes impossible to use the technical storytelling strategy for creating suspense that results from delaying the bifurcation (F.2). It also means that all the sequences have to be independent, sufficiently interesting

and, at the same time, conform to the general consistency of the story. They have to respect the narrative logic independently of its chronology – the moment in which they will be called to existence by the user.

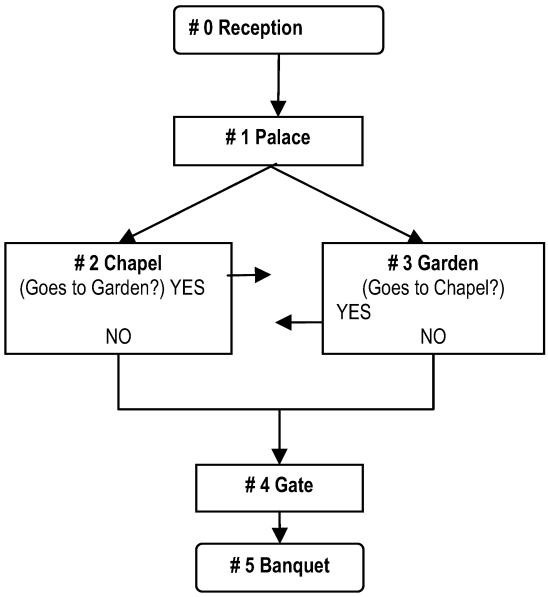


Figure 2. Model of the basic sequence organization

The places of the first (#0 Reception, #1 Palace) and the last sequences (#4 Gate, #5 Banquet) are necessarily pre-established for logistical purposes. However, the order of the intermediate sequences is random.

It is at this point – Sequences #2, #3, #n… – that the user's agency can be put into effect. In practical terms there is no way to escape the linearity of narrative, as each event has to occur in time, one after the other.

Also, in ***InStory*** (Ryan 2001) there can be found «purely selective interactivity» as the user can determine part of the plot, can swift perspectives by changing of avatar, and explore (with time) all the alternatives in the terrain gaining a global view of Regaleira gardens and palace (as a borgesian «garden of forking paths»); she can also retrieve documents/information from the server, play games and solve problems. There is some «productive interactivity», as the user participates in the «writing» of the narrative by choosing its path. Also, as the intermediate sequences are non-mandatory, the length of the story and the number of episodes experienced depend exclusively on the user's interest and resolve.

Anachronies – Recall and Foretelling

Basic to the distinction between «story» and «discourse» is the concept of «anachrony», where the rupture of the temporal order is used as a stylistic tool. Playing with time is one of the most important literary strategies, regarding the depiction of characters, psychology, the creating of an atmosphere, etc.

Structuralists (Genette 1996) have established three kinds of distinctions regarding the use of time: «order», «duration», and «frequency» of the events depicted. «Order» includes retrospection and anticipation – the narrative possibilities for recall / analepsis and foretelling / prolepsis. The other two modalities are naturally dependent of and subordinate to the former.

In traditional storytelling the interruptions of the natural chronological flow of the events happen often. These ruptures have a strategic function: either they delay the resolution of the bifurcation, or inscribe past and future tenses in the present of the story: recalling events that have already occurred, preparing actions to come, or delivering a revelation.

Considering that both these distinctions are made relatively to a moment that is considered a «now» (Shärfe and Øhrstrøm 2003), the present of each instant in which a story is being read or experimented: «The act of comprehending meaningful communication is possible due to our ability to perceive a "now" in relation to a past (retrospection) and in relation to a future (anticipation). Thus we comprehend a communicative structure by understanding new information as contingent upon previously perceived information, and by anticipating this or that future outcome of the communication. This is true of scientific as well as of narrative discourse, even though the means and pragmatic rationale may differ in terms of precision and granularity of intention. It is simply hard to describe the act of reading / seeing / listening / playing without the notion of a «now» from which past occurrences and future developments are understood».

It is relatively to this «now» that the elements occurring before and after are measured. The inscription of those moments is done through the use of verbal tenses (past/future) consistent with what the above authors call an A-logic (tense logic) in opposition to a B-logic (earlier /later): «In other words: retrospection and anticipation works by means of A-logical notions as long as the communication is proceeding. But it is equally obvious that when we are dealing with retrospective text comprehension, the perspective of the receiver may change from an inside view to an outside view, as we probe event's structures to grasp the signification and ramification of particular events. From this outside perspective, B-logical notions of "before" and "after" can be used just as natural as A-notions are used in describing comprehension based on empathy». This inside/outside perspective has been advocated (Ryan 2001) regarding the act of reading a text. Here, even if the user has beforehand knowledge of the Regaleira geographical map, she will not be able to perform these functions, as she is acting the story, with a first person character point of view. Also, in literary terms, the agency of those ruptures is usually attributed to a narrator, or a character with omniscient perspective.

Prolepsis (future-operator) has a foreshadowing effect, giving the author the possibility to taint plain occurrences with hidden and ominous narrative dimensions. An American Criticist (Frye 1957) avers that the use of such omens and foretelling are plot devices, independent of the actual belief in prophecies both for writer and audience.

Analepsis (past-operator) has a very practical function in what regards the structure of any story, as an instrument to delay the resolution of the bifurcation – the catalysis, in barthesian terminology (Barthes 1996). As implicit in the flashback it takes the story back in time, either to freshly narrate an event occurring before the actual «now», or to reiterate something already told that will acquire new meanings in the present situation. It is the «memory» internal to the narrative.

In a narrative like ***InStory*** there is no «memory». To exist, memory had to be registered as a past event to be recaptured in a certain «present». Here all moments are the «now», very close to a live reportage, much like a dramatic happening. Analepsis and prolepsis as a tool to promote the immersion of the reader– user do not exist, even if we consider, as before suggested, the all geographical space as a narrative map, a kind of gigantic prolepsis at the beginning of the story.

As it was said, there is not a pre-determined track to be taken by the user; and it is the user's route that creates the plot – the «discourse» – of the hunting story.

The order in which the user goes from one place to the other can be arbitrary, or not – for she may have a preconceived idea about the places she wants to visit. In any case, the narrative present will belong to the place where she is at the moment, each sequence will be the «now» that transforms every other node in «past»/«earlier» (if already visited) or in «future»/«later» (if to be visited).

Also, the user may receive a message informing that she has already visited a certain spot. If she insists in returning there, it becomes a different chapter in the narrative, as it receives the feedback of all the other episodes experienced until then. As each sequence of the story starts and ends at each spot, the other sequences – already visited, or to be visited – will embed themselves in the narrative plot as a different occurrence, enriched by the route already taken, enriched by the experiences already lived, gaining a surplus of connotation in the general bead necklace that this narrative «discourse» became.

Memory – of the past, or of the future – works in traditional narratives because it corresponds to the existence of events that are naturally implied in and by the story – by logic or verisimilitude; when the chronological order of the events exists, when it is possible to know the end of the story – where and when it is going to finish.

Prolepsis and analepsis do not function anymore relatively to the totality of the possible narrative to be told by Regaleira map, but only, and at a later stage, to the episodes associated with the spots that the user choose to visit. Yet, foretelling and recall can be recovered when the user, through the recording of her visit that will be made in the server, can recollect her experiences, and reorganize her own visit as a documentary. In a future phase the user will have the capacity to send texts or images, or even engage in dialogue/play with other users in the terrain. She will participate actively by creating new elements for the story enriching it with her own experiences rather than only deciding between given routes.

CONCLUSION

The implementation of the ***InStory*** project platform demanded a simple story to help guide the user who visited Quinta da Regaleira for the first time. It started as a very straightforward hunting narrative. Nevertheless, from it emerged quite an amount of speculative issues concerning the development of interactive narratives, in particular due to the physical constraints and technical limitations: the amount of information, the small size of a mobile screen, the fact that the users would be walking through an irregular geographical space.

These constraints became the foundations of its originality. The user is immersed in the story – as a character / narrator /avatar – inhabiting the «diegesis». The real time and space become a stage where the narrative takes place and another theoretical problem emerges: the literary ruse that narratology has called analepsis / flashback and prolepsis / flash-forward as we know them, fundamental for suspense, cannot be supported by a narrative of this kind.

The narrative structure imposed by the environment increased the potential for a fractal multiplication, as well as a new possible approach. Future work could include AI processes and programming, namely the use of genetic algorithms, rule preferences and revision to deepen the story; to find a solution for the lack of suspense problem; and also to introduce ethical rules in IF.

The fact that the user is physically and mentally active within the realm of the narrative, attests the unsuitability of traditional terminology to define new practices, and demands the reformulation of literary concepts or its adaptation to the new media, a new literary taxonomy.

REFERENCES

Barbas, H. and N. Correia. 2006. "Documenting InStory – Mobile Storytelling in a Cultural Heritage Environment". Ed. M. Zancanaro L. Bordoni, A. Krueger. First European *Workshop on Intelligent Technologies for Cultural Heritage Exploitation* (Aug.28) – *ECAI 2006, 17th European Conference on A.I.*. Riva del Garda, Italy. 6-12.

<http://helenabarbas.net/instoryEecaiHeritage_barbas_correia.pdf> [03.19.2009]

Barbas, H. 2006. *On Character – out of a critical Fallacy?* <http://helenabarbas.net/Character_HBarbas08.pdf> [03.19.2009].

Barthes, R. 1966. "Une Introduction à L'Analyse Structurelle du Récit". *Communications 8*. Seuil, Paris, France.

Benveniste, É. 1974. "L'Homme dans la Langue". *Problèmes de Linguistique Générale, 2*. Gall., Paris, France. 195-238.

Correia, N. et. al. 2005a. "InStory: A System for Mobile Information Access, Storytelling and Gaming Activities in Physical Spaces". *ACE2005-ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*. (Valencia, June 15-17). U. Politecnica de Valencia, Spain. <http://img.di.fct.unl.pt/InStory/publications/paper-ace2005.pdf> [03.19.2009]

Correia, N. et. al. 2005b. "Narrativas Interactivas em Dispositivos Móveis". *Nas Fronteiras do Imaginário. Artech 2005 – 2º. Workshop Luso-Galaico de Artes Digitais, XIII Bienal de Cerveira*. (Aug.27). V.N. de Cerveira, Portugal. 42-49.

<http://www2.fcsh.unl.pt/docentes/hbarbas/Textos/Narrativas_Interactivas_Artech_H_Barbas.pdf>

Davenport, G. 2003. "Dynamics of Creativity and Technological Innovation". *Digital Creativity*. Vol.15, nº.1, MIT Media Lab and Media Lab Europe, USA/Ireland. 21-31.

Davenport, G. 2005. "Desire Versus Destiny: The Question of Payoff in Narrative". *Caixa Forum MetaNarrative[s]? Conference*. (Jan.29). Barcelona, Spain. <http://ic.media.mit.edu/people/gid/papers/DesireVsDestiny2005.pdf> [03.19.2009]

Dell'Acqua, P. and L. M. Pereira. 2003. "Preferring and Updating in Logic-Based Agents". *Web-Knowledge Management and Decision Support*. INAP'01 – 14th. Int.Conf. on Applications of Prolog. Springer-Verlag, Berlin, N. York.

<http://centria.fct.unl.pt/~lmp/publications/online-papers/inaplnai03.ps.gz> [03.19.2009]

Douglass, J. 2007. *Command Lines: Aesthetics and Technique in Interactive Fiction and New Media* (PhD Diss.), Santa Barbara, Cal., U.S.A. < http://jeremydouglass.com/dissertation.html> [03.19.2009]

Frye, N. 1957. *Anatomy of Criticism: Four Essays*. Princeton U. P., Princeton, U.S.A.139.

Genette, G. 1966. "Fontières du Récit". *Communications 8*. Seuil. Paris, France.152-63.

Greimas, A. J. 1966. *Sémantique Structurale. Recherche de Méthode*. Larousse, Paris, France.

Holtzman, S. 1997. *Digital Mosaics: The Aesthetics of Cyberspace*. Simon & Schuster, N. Y., U.S.A.

Lambert, J. 2002. *Digital Storytelling: Capturing Lives, Creating Communities.* Life on the Water Inc., Berk., U.S.A.

Martins, T. et al. 2005. "InStory Client – a Browser for Spatial Narratives and Gaming Activities". *13º Encontro Português de Computação Gráfica*. (Vila Real, Oct.12-14) U. Trás-os-Montes e Alto Douro, Portugal. 133-138.

Mateas, M. and P. Sengers. 1999. "Narrative Intelligence". *AAAI 1999 Fall Symposium on Narrative Intelligence*. (Nov.5-7) American Association for Artificial Intelligence, Cape Cod, North Falmouth, Mass. <http://www.cs.cmu.edu/

afs/cs/user/michaelm/www/nidocs/MateasSengers.pdf> [03.19.2009]

Montfort, N. 2007. *Generating Narrative Variation in Interactive Fiction*. (PhD Diss.) University of Pennsylvania, Penn., U.S.A. <http://nickm.com/if/Generating_Narrative_Variation_in_Interactive_Fiction.pdf> [03.19.2009]

Nigay, L. and J. Coutaz. 1993. "A Design Space for Multimodal Systems – Concurrent Processing and Data Fusion". *INTERCHI '93 – Conference on Human Factors in Computing Systems*. Addison Wesley, Amsterdam, Neth. 172-8. <http://iihm.imag.fr/bouchet/ICARE/InterCHI93_DataFusion.pdf> [03.19.2009]

Pan, P. 2004. *Mobile Cinema*. (PhD Diss.) M.I.T., Mass. <http://alumni.media.mit.edu/~ppk/Publications/Pengkai%20Pan%20Thesis.pdf> [03.19.2009]

Pereira, L. M., 2008. "Darwinismo Literário e Computação". *Jornal de Ciências Cognitivas*, Jan. <http://jcienciascognitivas.home.sapo.pt/08-01_pereira.html> [03.19.2009]

Pereira, L.M. and G. Lopes, P. Dell'Acqua, 2009. "Inspecting and Preferring Abductive Models", to appear in: K. Nakamatsu, L.C. Jain (eds.), *Handbook on Reasoning-based Intelligent Systems*, World Scientific Publishers, 2009. <http://centria.fct.unl.pt/~lmp/publications/online-papers/rbis.pdf> [03.19.2009]

Prince, G. 1988. *Dictionary of Narratology*. Scholar Press, Aldershot, U.K. 21.

Propp, V. 1969. *Morphology of the Folktale*. Trans. L. Scott. U. Texas Press, Austin, U.S.A.

Ryan, M. L. 2001. *Narrative as Virtual Reality – Immersion and Interactivity in Literature and New Media*. The John Hopkins U. Press, Baltimore, U.S.A.

Shärfe, H. and P. Øhrstrøm. 2003. "Representing Time and Modality in Narratives with Conceptual Graphs". *Conceptual Structures for Knowledge Creation and Communication*. Springer-Verlag, Berlin, N. York. 201-14.

< http://www.hum.aau.dk/~scharfe/TimeMod.pdf> [03.19.2009]

Si, M. et al. 2006. "Social Norm Models in Thespian: Using Decision Theoretical Framework for Interactive Dramas". (Apr.) *AISB'06 Proceedings*. AISB'06 - SSAISA. University of Bristol, Bristol, U.K. 70-77.

Todorov, T. 1983. *Introduction to Poetics*. Trans. R. Howard. Harvester, Brighton, U.K.

Young, R. M. 2000. "Creating Interactive Narrative Structures: The Potential for AI Approaches". *The Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*. (March) <http://liquidnarrative.csc.ncsu.edu/pubs/potential.pdf> [03.19.2009]

Young, R. M. 2006. "Story and Discourse: A Bipartite Model of Narrative Generation in Virtual Worlds". (to be presented at) *Interaction Studies*. <http://liquidnarrative.csc.ncsu.edu/pubs/storyanddiscourse.pdf> [03.19.2009]

BIOGRAPHIES

HELENA BARBAS (1951) is Professor-Lecturer of the Departament of Portuguese Studies – F.C.S.H. – U.N.L., and a researcher of CENTRIA (http://centria.di.fct.unl.pt/). She holds a MA (1990) and a PhD (1998) in Comparative Literature – Literature and the Arts, and gained her "Habilitation" (2008) in Literature and Cyberarts.
Research interests: Multimedia Storytelling, IF, avatars, the usage of AI decision models and agents behaviour; serious games. She was a member of the *InStory Project* team. Presently she is preparing a project on serious games, *PlatoMundi*, aiming to introduce ethical issues in game playing.
<http://www2.fcsh.unl.pt/docentes/hbarbas/index.html>

NUNO CORREIA (PhD) is Professor at the New University of Lisbon – Computer Science Department, Faculty of Sciences and Technology. He holds a PhD and a "Habilitation" in Multimedia.
He is the coordinator of the Interactive Multimedia Group (IMG – http://img.di.fct.unl.pt). His research is centred on multimedia information processing, interaction and presentation. He has coordinated several projects in augmented and mixed reality, mobile storytelling, rich media spaces, personalization and collaborative video annotation, and ubiquitous learning.
< http://ctp.di.fct.unl.pt/~nmc/>

RECORD, REPLAY & REFLECT – A FRAMEWORK FOR UNDERSTANDING (SERIOUS) GAME PLAY

Anton Eliëns [1,2] and Zsófia Ruttkay [2]

[1] Intelligent Multimedia Group, VU University Amsterdam
[2] Creative Technology, University of Twente

KEYWORDS
serious games, learning, machinima

ABSTRACT

In this paper we set the first steps towards defining a framework for understanding game play based on recording user actions allowing for later replay, which we may characterize as symbolic machinima. After discussing issues of user tracking in virtual environments, we introduce a behavioral model for game play, discuss its relation to existing game reference models, and define a metrics to characterize behavioral discreprancy, for example from a norm scenario in a management game. Finally, we will indicate the possible benefits of our approach for understanding individual users' game play, as well as for the re-use of analysis data for authoring alternative game scenarios.

INTRODUCTION

Educators continuously face the problem of motivating their students to learn, to consult textbooks, do exercises, reflect on the material, and repeat this cycle over and over again. Although e-learning technology made the learning materials more readily available, as well as additional facilities for searching information, making rote-learning apparently obsolete, it did clearly not help in solving the motivational issues, that is the students' involvement with the subject, the time spent on studying and solving problems, and the unavoidable chore of repetition, that is inherent in all learning.

Recently, serious games have come to the rescue, allowing for a more active form of learning, Gee (2003), in a wide area of domains, including language learning, management, as well as behavioral attitudes, for example with respect to climate change, Eliens et al. (2007b). Yet, from an educators' perspective, such games may still fall short with respect to the essential reflective component of learning, since immersive in-game learning may in general be too implicit, in particular when no clear criteria (*right* or *wrong*) exist, as for example in more complex management games, that involve training of communication skills, Eliens & Chang (2007).

In this paper, we will introduce (the first steps towards) a framework for understanding game play based on symbolic machinima, that is the recording of user actions such that later replay is possible, either in a realistic format as in 'normal' machinima or in an adapted format, symbolically, to allow for reviewers comments or even the revision of players decisions. In summary, we envisage a number of possible application contexts for our approach:

- find exploration and learning strategies in serious games – and to allow feedback and review by human experts in a replay, think & reflect context
- provide user-tailored services – e.g. museum guide, or personalisation and recommendation scenarios
- learn users multimodal behavior – for control or as sample for NPCs (Non-Player Characters)

Although we are in the process of implementing our approach in the XIMPEL platform[1], we see our contribution on a high and general level, which may be applied for a variety of different platforms. In our own research, not only do we wish to use the developed technology in the context of serious games based on interactive video, but also in explorative educational environments for (virtual) musea. Another application domain is training (such as a virtual fitness trainer), where feedback strategies are of major interest, Ruttkay et al. (2006).

structure The structure of this paper is as follows. First we will briefly characterize serious games, in particular in distinction to e-learning and online information facilities. Then we will discuss the background and inspiration of our approach and characterize the notion of symbolic machinima. After dealing with issues in tracking user actions in virtual environments, we sketch the outlines of a behavioral model for game play, and discuss its relation to existing game reference models. We then define a metrics for behavioral discrepancy, which allows for relating user behavior to possible norm scenarios, and conclude by giving a brief overview of the benefits of our approach with respect to understanding individual users' game play as well as the use of analysis data for authoring alternative game scenarios.

[1] ximpel.net

SERIOUS GAME(S) RE-CONSIDERED

The literature on (serious) games abounds. To set the stage, we limit ourselves to a single quote from *virtual heroes*[2], that we more extensively discussed in Eliens & Chang (2007):

> *Serious games and simulations are poised for a second revolution. Today's children, our workforce and scientists are increasingly playing, learning, and inventing in visually intensive "virtual" environments. In our increasingly experiential economy, immersive educational and training solutions are needed to advance the workforce of tomorrow. Game-based learning and technologies meet this challenge.*

Yet, as we indicated in Eliens & Chang (2007), from the observation that serious game technology meets current educational challenges, it is still a long way to actually develop interesting serious games, that can not only compete with e-learning facilities in addressing educational goals, but may also be considered to be sufficiently playful to count as game(s) whatsoever.
A very helpful set of criteria for distinguishing games from other (online) applications were presented to us in a workshop on educational games in a museum context[3], which mentions 4 essential characteristics to assess the extent whether an application may considered to be a game:

- *challenge* – relevance, feedback, confidence
- *curiosity* – cognitive or sensitive discrepancy
- *control* – contingency, choice, power
- *context* – intrinsic or extrinsic metaphor(s)

Leaving a more detailed interpretation of the characteristics *challenge*, *curiosity* and *control* to the inventiveness of the user, who may use the keywords for support, the *context* characteristic, however, needs some elaboration. As the discussion in the workshop indicated, it is rather easy to use extrinsic metaphors or game formats for arbitrary content. For example, a memory game can be reused over and over again, just by changing the images according to the topic or subject, that is language learning, climate change, etcetera. These kind of mini-games or casual games lend themselves to a variety of learning tasks and may be constructed using pre-defined game formats. More difficult, however, is to construct games with an intrinsic relation to the topic, and in the workshop for educational games in a museum the best suggestion was a scenario that imprisoned the player in the museum by night, haunted by the figures depicted on the paintings ...

[2]www.virtualheroes.com
[3]www.archimuse.com/mw2008/abstracts/prg_335001733.html

Apart from *context*, *challenge* and *control* seem to be the major parameters for modelling user actions in terms of, respectively, goals or topics and strategy and choices, or in other words scenarios with more or less well-established courses of behavior.

REALIZING SYMBOLIC MACHINIMA

The background and (partial) inspiration of our work is formed by the FP7-ICT-2007-3 Project CAESAR (Computer Aided Experiential Story Acquisition and Reuse), initiated by Pedro Gonzalez of the University of Madrid. Key elements in the proposal are *semantic-enabled machinima* and *end-user narrative content creation*. Not wishing to judge the wisdom of those who had to review the manifold of proposals, and notwithstanding the necessarily heterogeneous content of such proposals, imposed by the constraints of enforced cooperation between widely different international groups, we are nevertheless disappointed by the rejection of our request for funding, and decided to continue this valuable line of research within the proximity of our own groups.
Paraphrasing the CAESAR project proposal, we may observe that: The idea of producing animation movies using the tools and resources available in a game and rendering them with a 3D game engine, appeared in the early 90s and is now known under the term Machinima (machine cinema). Interest in machinima is growing as demonstrated by sites such as Tube2SL (tube2sl.com), a Second Life based Broadcasting Network for machinima productions, or WeGame (wegame.com), a media sharing platform for gamers, in public beta since January 2008, where gamers can post and share their in game recordings. New top of the line commercial games include machinima tools, such as Halo 3, Microsofts XBox 360 flagship game title released at the end of 2007, whose Saved Films feature is one of the main innovations with respect to Halo 2.
Obviously there are many benefits to be expected when extrapolating machinima to semantic-enabled or symbolic machinima. Quoting the CEASAR proposal:

- content creation – With the rise of persistent worlds, content designers for video games are more in demand than ever. Roleplaying games have always needed truckloads of content, and now that these worlds have been made seemingly endless, the need for more content is ongoing. Game content generation by example would provide a highly cost-effective solution to this problem.
- community building – Community building is a key issue for games and even more for the success of virtual worlds. CAESAR would provide new and more powerful ways for the community members to share their experiences in the virtual world in a way that is not possible in the real world, exploiting full observability

and the possibility of living virtual experiences wearing somebody elses shoes.

- demonstration material – Animation movies have applications for entertainment, communication and education but their production is too expensive for most potential users. Semantic enabled machinima would open the range of users of animation movies with affordable cost.

Our specific role as partner in the CEASAR project was to address the issue of learning about user behaviour, in order to provide sensible feedback, from semantically logged interactions. Tracking representations are envisaged at a high level for which we would like to develop a general framework, which can be used as a reference for developing concrete methods for specific (replay/feedback/authoring) purposes in dedicated application contexts. All what is assumed is that user and system actions are logged with time stamps and semantic labels of some kind. Below we introduce the different aspects of the analysis and learning framework, as it pertains to

level(s) of behaviour

- bodily aspects – to learn communicative multimodal behaviour patterns, or new gestures or motion sequences. In this case the tags are low-level bodily signals (blink, nod, beat), which may be gathered by some sensory input devices (vision, pressure sensor).
- cognitive aspects – to analyse or learn strategies, action patterns of the user. The labels are on a high semantical level, indicating choices made by the user in given situations.
- narrative/rethoric aspects – analyse dramatic effects, features about the emerging narrative

USER TRACKING IN VIRTUAL ENVIRONMENTS

Many of the technologies to realize games or rich-media interactive applications, including X3D/VRML, Flex/AS3, and the Halflife 2 SDK, as well as Second Life, use events to capture user actions, which in its turn may be stored and programmatically invoked to re-create or simulate a sequence of user actions. In Eliens (2000), moreover, we have demonstrated how to use object-technology to create event-driven simulations capturing complex state information, allowing for complete undo and redo actions.
As reported in Eliens et al. (2007a), we used event-capture techniques to create guided tours in virtual environments for cultural heritage by tracking expert behavior, even allowing for the user of guided tours to express preferences for particular choices by (implicitly) defining weights on the influence of experts deciding on alternative choices. Thus having a database of tours from a number of experts, we may offer the user a choice of tours, and even allow to give priority to one or more of his/her favorite experts, again simply by adjusting the weighting scheme.
As more fully explained in Eliens & Wang (2007), guided tours, in the digital dossier, may take one of the following forms:

- automated (viewpoint) navigation in virtual space,
- an animation explaining, for example, the construction of an artwork, or
- the (narrative) presentation of a sequence of concept nodes.

In practice, a guided tour may be constructed as a combination of these elements, interweaving, for example, the explanation of concepts, or biographic material of the artist, with the demonstration of the positioning of an artwork in an exhibition space.
As a pre-condition for the construction of guided tours based on user tracking we identified the requirement that navigation consists of a small number of discrete steps. This excludes, at first sight, the construction of arbitrary guided tours in virtual space, since it is not immediately obvious how navigation in virtual space may be properly discretized. As an additional requirement, it must be possible to normalize interaction sequences, to eliminate the influence of short-cuts, and to allow for comparison between a collection of recordings. The application of the techniques developed for constructing guided tours requires that choices are discrete and only apply to capture navigation in virtual environments when we find find proper ways to encode such navigation as a small finite collection of discrete steps. Also in the discrete case, however, we must be able to normalize navigation paths, in order to compare and weigh the navigation sequences of multiple users. For the actual playback, as a guided tour or replay, a decision mechanism may be needed that finds the proper advice or weight at each decision point to select the optimal step, according to some decision rule that takes the weighting scheme as for example expressed in a norm-scenario into account.
In Eliens et al. (2007c), we have indicated how tracking user behavior may be realized in Second Life using an elementary web-server containing the following resources:

web server

- /seen?user=SomeAvatar – records the presence of SomeAvatar
- /touched?user=SomeAvatar – invokes object API for user SomeAvatar
- /set_tag?user=SomeAvatar&tag=FavoriteTag – records SomeAvatar's favourite tag

For example, in response to a 'touch' event, invoking *touch* results in consulting the database for the user's tag and possibly sending a request to the object API performing some action on behalf of the user or recording a user's favorite tag. These invocations could easily be extended with time tags to enforce linear ordering.

TOWARDS A BEHAVIORAL MODEL FOR GAMEPLAY

Storing events resulting from user actions, possibly together with events influencing the game state autonomously, generated by the game system, gives us an immediate, albeit low level, way to record game play, allowing for machinima-like replay. However, in order to be able to provide meaningful feedback on the choices made by the user during play, we need a more high level representation of the users' behavior and choices made when the user is confronted with particular challenges. For inspiration, we first looked at what *game interaction patterns* might have to offer, Björk & Holopainen (2005), but apart from the *Score* pattern, little support was found for symbolically representing user actions, due to the rather abstract nature of patterns.

To simplify matters, we decided to reduce the representation problem to modelling the behavior of users at choice points in interactive video, as supported by the XIMPEL platform, and how particular choices reflect the attitudes or preferences of users with regard to particular topics. Although limiting ourselves to interactive video may seem to be too restrictive, as we argue in Eliens et al. (2008) interactive video may provide an excellent basis for game play, and as for example demonstrated in our Dante-inspired *Journey to Hell*[4] application, allows for assessing what we may call in this case a *moral profile* of the user, simply by recording the choices made by the user on questions of a moral nature. In the hope of being able to extend the model to more rich forms of game play, this approach allows us to take a model originally meant to capture ratings and recommendations, as explored in Eliens & Wang (2007), and extend this to represent attitudes and preferences with regard to topics of interest. See also Van Setten (2005) for information on recommendation and user modeling.

Since XIMPEL was originally developed for a climate game, Eliens et al. (2007b), we will take climate issues and attitudes towards measures affecting global warming or the effects thereof as a starting point to illustrate our approach, which we will present without going into very much formal detail. As an example, let's look at how we may model the behavior B of a user in the context C of a debate between experts, where the user is challenged to take action to provent flooding of the Netherlands due to global warming, for example by reducing the emission of CO2. In outline we may represent this situation as:

B = [choice = measures, action = reduction]
C = [context = debate, challenge = flooding]
P = [control(human_influence) = true]

[4] www.cs.vu.nl/∼eliens/ximpel/dante

Here we represent *behavior* by relating actions to choice points, *context* by making the situation explicit in which the choice is presented as well as the challenge the user is confronted with, and finally *preferences* by indicatiog how the user takes control. Admittedly, for the derivation of preferences based on behavior in context we would need a rather strong ontology describing the semantic relations within the game domain. Nevertheless, although still a far cry from a formal model, having a suitable representation for *choices*, *actions* as well as the features defining *context*, *challenges* and *preferences* would allow us to record game events on a sufficiently high level, so that they may later be used for meaningful feedback.

A REFERENCE MODEL FOR EFFECTIVE GAME PLAY

In Eliens & Chang (2007) we introduced a reference model for game play, to be able to decide on the effectiveness of the players' strategies and actions in attaining the goals set in service managements games. The basic model, adapted from Juul (2005), consisted of the following elements:

reference model

- *rules* – service management protocols
- *outcome* – learning process
- *value* – intellectual satisfaction
- *effort* – study procedures
- *attachment* – corporate identity
- *consequences* – job qualification

Relating this model to our *challenge*, *curiosity*, *control* and *context* criteria, we may regard *rules* and *effort* as constitutive factors for *challenge*, *outcome* and *effort* as belonging to *control*, and *attachment* and *consequences* as belonging to *context*, to which, naturally, also *rules* bear a strong relation.

For service management games, we added two more criteria to the model, namely *scenarios* and *reward*, dealing with the (serious) content of the game:

- *scenarios* - problem solving service management
- *reward* - service level agreement

Both the notions of *scenario* and *reward* are essential in understanding (serious) game play, since they allow to indicate a specific level of attainment to which the player must comply, in order to be considered to have played the game effectively.

From a different perspective, in terms of the notions introduced in a behavioral model of game play, as introduced in the previous section, we may classify *rules* as belonging to *context*, *outcome*, *consequence* and *value* to *preferences*, and *effort* to *behavior*. Adding *scenarios* and *rewards* helps in defining *challenges*,

and, in principle, to define *norm scenarios*, setting a standard for the most appropriate actions, which is clearly relevant for serious games intending to bring about an attitude change, for example in behavior affecting climate change.

METRICS FOR BEHAVIORAL DISCREPANCY

Given the notions of *scenario* and *reward*, as introduced in the previous section, we cannot resist to speculate on how we can define *norm* scenario(s) and associated metrics to assess behavioral discrepancy, that is the degree in which the user deviates from a desired course of action, a particular position or set of preferences. To allow for a more formal treatment, it seems most convenient to adapt the behavioral model introduced earlier, by reducing behaviors to consist of actions only, taking context into account implicitly, and to redefine preferences as rewards, which may conveniently be expressed as scores over predefined result parameters, such as (*people*, *planet*, *profit*) in the case of our climate game.
Representing the combined result parameters (*people*, *planet*, *profit*) as vectors of features characterizing preferences for aspects of the individual parameters allows for defining a metric over the space of preferences defined by the result parameters, using a standard distance metric, as we originally did for recommendations in Eliens & Wang (2007).
Using such a metric allows us to assign a rating or an indication of relevance to the result parameters, as illustrated by the following example. If we assume that alternative actions have effects as listed below

a_1 = [planet = green, profit = high]
a_2 = [planet = green, people = happy]
a_3 = [planet = red, profit = high]
a_4 = [planet = red, people = happy]

then we may, in an abstract fashion, deduce that if $d(a_1, a_2) < d(a_1, a_3)$ then $r(profit) < r(planet)$, for a rating funtion r. However, if $d(a_1, a_3) < d(a_1, a_2)$ the reverse is true, that is then $r(planet) < r(profit)$. In other words, actions involving only particular features of any of the result parameters may influence the final result when taking a particular position or preference as the norm.
Given a metric on preferences, which induces a metric on actions, and a norm scneario, with a recommended sequence of situations $\ldots, s_{n-1}, s_n, \ldots$, with for s_n possibly alternative actions $a_1, a_2, \ldots$, we may adapt the (implied) preference of the user, when the user chooses to select alternative a_k instead of accepting s_n as recommended by the norm scenario, to adjust the score by taking into account an additional constraint on the derived score. Differently put, when we denote by $s_{n-1} \mapsto s_n/[a_1, a_2, \ldots]$ the presentation of issue s_n with as possible alternative actions $a_1, a_2, \ldots$, we know that $d(s_{n-1}, a_k) < d(s_{n-1}, s_n)$ for some k, if the user chooses for a_k.
Admittedly, apart from easily skipping over representational issues, we have omitted many of the necessary formal details. We refer to Cesa-Bianchi and Lugosi (2006) for readers wishing to explore the mathematical details of our approach.

BENEFITS OF A QUALITATIVE APPROACH TO REFLECTION AND FEEDBACK

Whereas quantitative results, as for example obtained in tests or exercises in specific skills, may be worthwhile in domains such as language learning or, for that matter, the operation of vehicles, a more qualitative approach seems necessary for (serious) game tasks that involve communication skills or strive to induce attitude changes, as is the case with management games or games related to topics of societal interest, such as climate change and security.
With respect to individual users, an approach as sketched in this paper offers the opportunity to analyse behavioral patterns of a single user interacting with the system. The issues involved here, or the potential usage of such analysis include:

- one particular user vs. group behavior
- one session or multiple sessions
- novice vs. experienced user or expert
- possible recommendations or advice

Our approach, which we have summarized in the title of this paper as *record, replay & reflect* originated from the wish to provide feedback and replay, preferably in a user-friendly textual format, that is to present segments of the interaction for viewing interesting/problematic parts, give summary about the interaction and performance (e.g. in a learning environment), either for the user him/herself or as a summary for system developers providing feedback about the usage of the facilities within the system. Other goals for which our approach may be used, as expressed in the original CAESAR proposal, include:

goal(s) of analysis

- enhance the behaviour repertoire of user-control or of virtual characters – in the game, based on analysis of user behaviour and interpreting it on a higher level (e.g. new types of greeting signals, new escape behaviour).
- author exploration paths and navigation strategies – use the interaction sample (possibly by an expert) as a reference e.g. for selective tours in a virtual environment or learning applications.

- learn user profile – elicit typical interaction or behaviour characteristics of a given user or user group, and use this information to re-design or dynamically adapt the system.

Finally, as also mentioned in the CAESAR proposal, where the deployment of semantic-enabled machina for *content authoring* played a central role, we envisage the potential reuse of game play in different contexts or platforms. With a sufficiently high level representation in a suitable interchange format such as XML, we would also like to explore the reuse of missions and scenarios in different contexts, and even different platforms, similar as the proposed Collada[5] standard does for (graphics and physics) game content, as a means to accommodate the authoring of narratives and story lines.

CONCLUSIONS

In this paper we have sketched the outline of a framework for understanding game play, which may be used for providing meaningful feedback to (serious) game players, allowing for replay on a sufficiently abstract level by deplaying semantic-enabled machinima.
Although we have partially implemented aspects of our approach in the XIMPEL platform, addititional research is needed to arrive at a sufficiently complete representation scheme for capturing events, user actions and resulting game state changes. It seems that, in particular, we must pay more attention to the domain ontology underlying a game, to enable the construction of user profiles using inferential reasoning based on the actions taken by the player when confronted with choices or challenges in the pursuit of a mission or scenario.

Acknowledgement(s) Thanks are due to the guys from the Clima Furura Labs, Marek van de Watering, Hugo Huurdeman and Winoe Bhikharie, who among other things developed the XIMPEL platform. We also wish to express our thanks to Pedro Gonzalez of the University of Madrid who initiated and got us involved in the CEASAR project.

REFERENCES

Björk S. and Holopainen J. (2005), *Patterns in Game Design*, Charles River Media

Cesa-Bianchi N. and Lugosi G. (2006), *Prediction, Learning, and Games*, Cambridge University Press

Eliens A. (2000), *Principles of Object-Oriented Software Development*, Addison-Wesley Longman, 2nd edn.

Eliens A., Wang Y. van Riel C. and Scholte T. (2007), 3D Digital Dossiers – a new way of presenting cultural heritage on the Web, In Proc. Web3D 2007, ACM SIGGRAPH, pp. 157-160

Eliens A., van de Watering M., Huurdeman H., Bhikharie S.V., Lemmers H., Vellinga P. (2007), Clima Futura @ VU – communicating (unconvenient) science, In Proc. GAME-ON 07, Nov 20-22, University of Bologna, Marco Roccetti (ed.), pp. 125-129

Eliens A. Feldberg F., Konijn E., Compter E. (2007) , Mashups in Second Life @ VU, In Proc. GAME-ON 07, Nov 20-22, University of Bologna, Marco Roccetti (ed.), pp. 130-134

Eliens A. and Chang T. (2007), *Let's be serious – ICT is not a (simple) game*, In Proc. *FUBUTEC 2007*, April 2007, Delft

Eliens A. and Wang Y. (2007), Expert advice and regret for serial recommenders, In Proc. EUROMEDIA 2007, L. Rothkrantz and Ch. van der Mast (eds), Eurosis, pp. 111-118

Eliens A., Huurdeman H., van de Watering M., Bhikharie S.V. (2008), XIMPEL Interactive Video – between narrative(s) and game play, In Proc. GAME-ON 2008, Valencia, Spain

Juul J. (2005), *Half Real – Video Games between Real Rules and Fictional Worlds*, MIT Press

Gee J.P. (2003), *What video games have to teach us about learning and literacy*, Palgrave Macmillan

Ruttkay Zs., Zwiers J., van Welbergen H. and Reidsma D., Towards a Reactive Virtual Trainer, Proc. of IVA 2006, LNAI 4133, pp. 292-303. 2006.

Van Setten M. (2005), *Supporting People in Finding Information – Hybrid recommender Systems and Goal-based Structuring*, Ph.D. Thesis, Telematica Institute Netherlands

Anton Eliens (PhD) is lecturer and coordinator of multimedia at the VU University Amsterdam, and was recently appointed at the University Twente as professor creative technology / new media. He has been closely collaborating with Zsófia Ruttkay and is experienced in web-based interactive media such as Second Life, interactive video, and the application of such technologies in serious games like Clima Futura.

Zsófia Ruttkay (PhD) is Assoc. Prof. at the University of Twente and leads the Creative Technology Working Group. She has expertise is in creating styled multimodal behaviour and communication strategies for virtual humans. She also has a strong interest in educational games, and has been developing an interactive virtual trainer application.

[5]en.wikipedia.org/wiki/COLLADA

ENHANCED AUDIO VISUAL RESEARCH

A VERTICAL STEREOSCOPIC SYSTEM BASED ON 1D IMAGE MATCHING

Viorica Patraucean
IRIT-ENSEEIHT, VORTEX Team
email: vpatrauc@etu.enseeiht.fr

Jean Conter
IRIT-ENSEEIHT, VORTEX Team
email: conter@enseeiht.fr

KEYWORDS
stereovision, catadioptric, image matching, Viterbi, gauss filter

ABSTRACT

Stereoscopic systems have proved their large applicability the last years in various domains: robotics, surveillance, 3D maps etc. Conceived as two-camera systems or as catadioptric mono-camera systems, these equipments rely on the possibility of gaining the third dimension of a scene by using two images of it, taken from different points of view. In this paper, we present an original stereoscopic system, as well as the dedicated image matching process. In order to respect timing constraints, we base the image matching on 1D processing. We follow two steps in achieving the 3D reconstruction: first, we detect and match interest points, by using progressive Gaussian filtering and correlation measures, then we proceed to global matching, using the first step in order to improve accuracy. The results were validated using poor quality real images, taken from a cave, as part of a virtual visiting project.

1 INTRODUCTION

Stereovision has greatly preoccupied the scientific branch due to its possible applications: autonomous mobile robots, 3D maps, surveillance systems etc. The most common paradigm in stereovision consists in using two (or more) images of a scene, taken from different points of view, in order to accomplish its 3D reconstruction. For example, considering this statement, we could image ourselves virtually visiting turist attractions that are hundred miles away and still having the real 3D perception of the relief of that place. To accomplish such a scenario, an adequate image acquisition system should be provided, as well as the necessary image processing background.

As for the image acquisition, several approaches have been proposed: the classical systems use two cameras, horizontally laid out - copy of the human visual system. In this manner, they are able to capture two images simultaneously, from slightly different points of view. More sophisticated systems use a single camera and a mirror system (Nayar 1988, Southwell et al. 1996, Nayar and Baker 1998), and are called catadioptric systems. Due to the mirrors' position, it is possible to have two projections of the scene in only one image. As a comparison, the classical systems have to deal with distortions that are easier to manage in a one-camera system. On the other hand, the use of a mirror system requires a meticulous adjustment and calibration and basically a more complicated geometry to tackle with.

In order to extract the relief information from the stereoscopic images, two steps are to be followed: first, the images must be matched (Step 1), i.e. we have to find the projections of one 3D point in the two images and afterwards, using a triangulation scheme, the full 3D coordinates including the distance between the camera and the 3D point are computed (Step 2).

Image matching has been widely studied (Brown et al. 2003, Harris and Stephens 1988, Lowe 2004, Cox et al. 1996, Criminisi et al. 2003), but still no general solution has been found, as there is a tight dependence with the camera used. There is no algorithm that could be applied from a camera to another as it is; adjustments need to be made in order to take into consideration the new conditions of lighting, the noise level etc.

The stereoscopic system that we propose, called PCam-ST, has a unique linear CCD (efficient for obtaining high resolution) and a mirror system, vertically disposed. An image contains thus two projections of the scene, one above the other. The dedicated image matching solution follows two steps: it first detects and matches interest points and then extends the matching process to the whole image, using the interest points as pivots, in order to achieve better accuracy.

Notations and background

When talking about classical systems, the two stereoscopic images are called "left image" (I_l) and "right image"(I_r) respectively, and the two projections of a 3D point are refered to as x_l and x_r. In our case, the scene perspectives being obtained one above the other in the same image, we talk about "top image" (I_t) and "bottom image"(I_b) and we denote the projections of a 3D point as x_t and x_b. It is obvious that, in lateral stereoscopy, the correspondent pixels are located on the same line (the epipolair constraint is used if not). In an image issued by PCam-ST the correspondences need to be searched on columns.

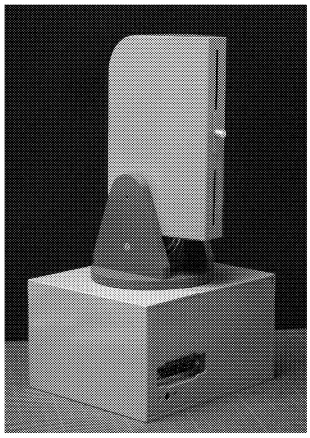


Figure 1: *PCam-ST –a panoramic stereoscopic camera*

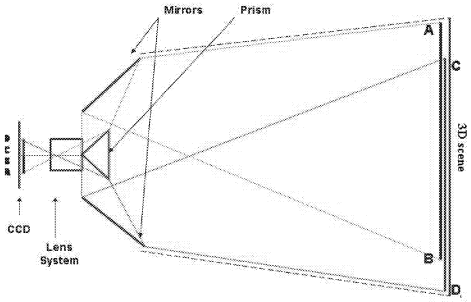


Figure 2: *PCam-ST Structure*

2 THE PCam-ST STEREOSCOPIC SYSTEM

PCam-ST is an *omni directional stereoscopic catadioptric system, with a unique linear CCD and a single optical system*(Figure 1). Although the camera uses a single CCD, it is still possible to simultaneously obtain the two stereoscopic images, thanks to the mirror system (Figure 2). Due to it, a single image issued by PCam-ST contains two projections of the scene, one above the other.

The mobile head of the camera, able to rotate about two axes, allows the complete coverage of the surroundings. The advantages of using a unique CCD and a single optical system over a classical system reside in terms of unified distortions, simplified calibration – a single system needs to be calibrated, and reduced cost – a single camera needs to be acquired.

Another important particularity of our system is the use of a linear CCD. It consists of three arrays of pixels (one for each colour channel), of 2700 pixels each. One pixel is 8x8μm and the distance between two arrays is 32μm.

The CCD "scans" the scene, the image being obtained column by column. Even if the process is slowed down by the chosen solution, the motivation of using this type of CCD is given by the elevated resolution which can be obtained. A classical matrix CCD with the same resolution is suspicious to have an important number of faulty pixels.

The Step 2 of the 3D reconstruction depends on the values of the intrinsic parameters of the stereoscopic system. We have obtained these values using a theoretical model for PCam-ST (Figure 3), which ignores the mirrors and considers the two halves (top half and bottom half) of the CCD as two theoretical CCDs, vertically disposed. It is important to mention that the top image (I_t) is obtained on the lower part of the CCD and the bottom image(I_b) is obtained on the upper half of the CCD. The camera calibration, which we don't describe here, provided the values of the intrinsic parameters: the positions of the invariant points and the distance between the two theoretical axes (E=94.7mm).

Distance Computation

For the classical systems (Figure 4), the distance between the camera and the scene points is given by:

$$d = \frac{f * E}{x_l + x_r} \tag{1}$$

where E is the distance between the two parallel optical axes and f is the focal length.

In our case, we are able to determine the value of the distance using the theoretical model (Figure 3). The distance is given by:

$$d = \frac{E\sqrt{1 + tan^2(\alpha_b)}}{tan(\alpha_t) + tan(\alpha_b)} \tag{2}$$

where α_t and α_b are the angles between the theoretical axes and the projection rays of the 3D point M.
The angles values are encoded in a table, which associates to each pixel the corresponding angle.

3 1D IMAGE MATCHING

Image processing, as we know it, is, by excellence, a 2D domain. However, the fact that our image data is retrieved column by column from PCam-ST encouraged us to take into consideration a 1D matching method, which highly reduces the computational complexity. We have conceived the image matching as a two–step process: first, we match the interest points located on a column and then we extend the matching to the whole column, using an algorithm based on dynamic programming. The implemented algorithms take into account the ordering constraint, and also use an a priori knowledge on the scene, by imposing a maximum disparity.

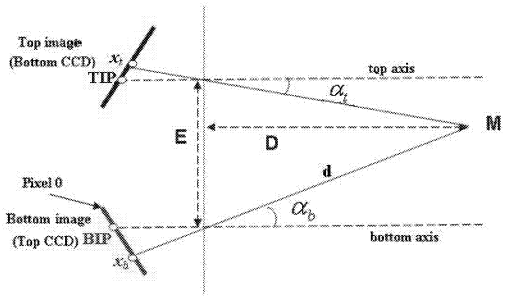


Figure 3: *PCam-ST Theoretical model*

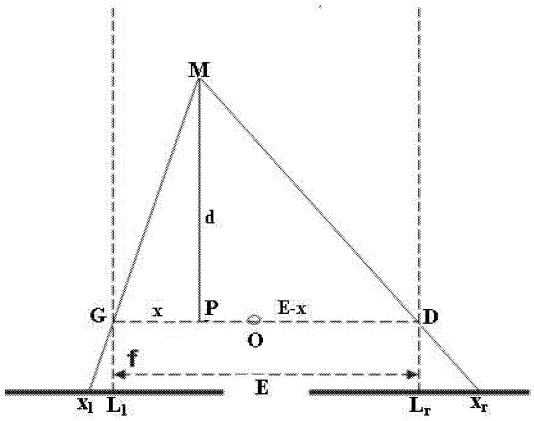


Figure 4: *Computing the distance for a classical stereoscopic system*

3.1 1D Interest Points Matching

There are several well-known algorithms for detecting and matching interest points in an image (Harris and Stephens 1988, Lowe 2004), but we are interested in using a 1D approach. Therefore, we propose an algorithm based on Gaussian filtering (for interest point detection)

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{3}$$

where σ is the standard deviation, and zero normalized cross-correlation measure $ZNCC$ (for matching)

$$ZNCC(x_t, x_b) = \frac{\sum_{u \in W} (w_t(u) - w_{tm})(w_b(x+u) - w_{bm})}{\sqrt{\sum_{u \in W} (w_t(u) - w_{tm})^2 \sum_{u \in W} (w_b(x+u) - w_{bm})^2}} \tag{4}$$

We have applied consecutively a 1D Gaussian filter (3) with the deviations $\sigma = 1, 3, 5$ respectively, and we have identified the points that kept their property of local maximum in the three cases. The identified points were then matched using the ZNCC measure (4), as it provides invariance to lighting changes. The 1D window size was of 11 pixels, centred on the pixel of interest. The correlation score for a pair of pixels to be accepted as matched was 0.8.
The tests carried out on a set of 9 noisy images showed a medium of 60% of correct matches.

3.2 1D Dense Matching

Although the interest points matching can be useful in certain applications, it proves to be insufficient for others: the key reason is that the results are sparse. Therefore, a dense matching is required, which usually implies a high computational complexity, an important inconvenient in applications with reduced resources.

A method widely used in dense matching due to its timing performance is the dynamic programming, and more precisely, the Viterbi algorithm (Viterbi 1967). According to this approach, the image matching can be expressed in terms of an optimal path research into a graph. The optimal path is obtained by minimising a cost function, as each segment of the path implies a cost.

Cox (Cox et al. 1996) and Geiger (Geiger et al. 1995) proposed two possible graph structures for the research space (Figure 5). The difference is stated by the connectivity of the graph. We have used the Cox structure, as it has reduced complexity, the price in terms of achieved accuracy being insignificant. The maximum disparity (200 pixels) was taken into account to reduce the research space. The use of this structure ensures the respect of the ordering constraint, as the possible moves from the pixel (i, u) are towards pixels: $(i+1, u+1)$ when pixels $(i{+}1)$ and $(u{+}1)$ are considered as matched, $(i+1, u)$, $(i, u+1)$ when the pixel u, respectively i is occluded.

The associated costs to each movement can be expressed (cf. (Criminisi et al. 2003))as:

$$M(i,u) = min \left\{ \begin{array}{l} M(i-1,u) + C_o \\ M(i-1,u-1) + C(i,u) \\ M(i,u-1) + C_o \end{array} \right\} \tag{5}$$

where C_o is the penalty (constant value) for accepting an occlusion and $C(i,u)$ is the matching cost, given by the sum of the squared differences SSD correlation measure

$$SSD(i,u) = \sum_{u \in W} (w_t(u) - w_b(i+u))^2 \tag{6}$$

The SSD measure was applied on linear windows, of 7 pixels, centred on the current pixel.

The algorithm uses two matrices of the same dimension: the first one, M, keeps the costs computed accord-

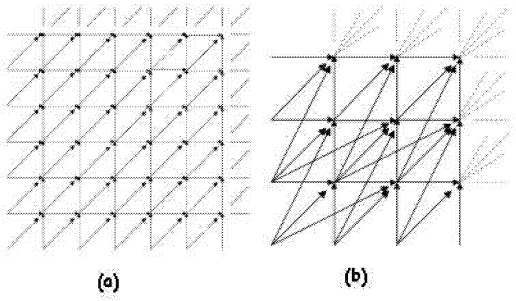


Figure 5: *Graph structures proposed by a)Cox and b)Geiger*

ing to (5), and the second one W, keeps the track of the direction that generated the minimum cost. The matrix W is used at the end to find the optimal path, which is obtained by scanning W in reverse order.

The results obtained using the dynamic programming are satisfactory, even for poor-quality images (Figure 6, Figure 7). We want to emphasize the fact that the algorithm was carried out on the original images, without any filter being applied, in order to keep intact the image data. The mean error is $\epsilon_m = 5\%$ (i.e. an error of 5cm for a distance of 100cm). Still, the error significantly grows ($\epsilon_m \sim 20\%$) in the neighbourhood of large occlusion regions.

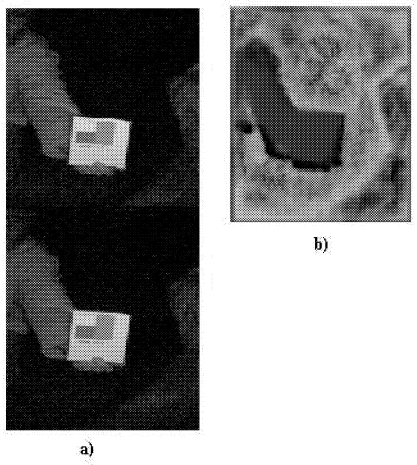


Figure 6: *a)Original (sub)image (1698x110) containing the two projections of the scene, taken in a cave, with an important noise level. b) Depth map obtained using 1D Viterbi algorithm. The distance is proportional to the grey level. The results in the dark region of a) are irrelevant.*

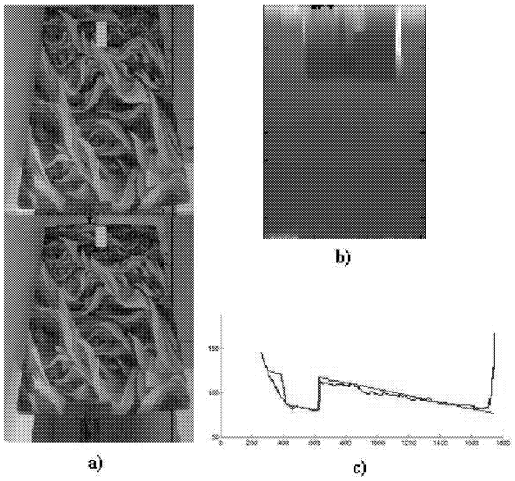


Figure 7: *a) Original image of a white square, placed in front of a leaning poster. b)Zoom on the depth map (white square region) obtained using 1D Viterbi algorithm. c) Example of the distances found for the 300th column of the original image – blue trace. The red trace represents the ground truth. The error is more significant if the occlusion region is at the beginning/end of a column.*

3.3 Guided Dense Matching

The problems encountered by the Viterbi algorithm in large occlusion regions led us to the idea of using the interest points previously detected to guide the optimal path through the research space. Torr et Criminisi have proposed this idea in (Torr and Criminisi 2004). The cost function needs to be adjusted, in order to take into consideration the matches found by the interest point detector. First, we have to evaluate the error probability of the detector λ_e, by running a set of tests. Afterwards, we can compute the probability $P(i,u)$ for the pixel i from the top column to be correctly matched with the pixel u from the bottom column as follows:

$$P(i,u) = \begin{cases} 1-\lambda_e & \text{if } (i,u) \text{ is a detected match} \\ \frac{\epsilon\lambda_e}{m-1} & i \text{ occluded} \\ \frac{(1-\epsilon)\lambda_e}{m-1} & \text{otherwise} \end{cases} \quad (7)$$

where ϵ is the probability of an occlusion occurrence and m is the pixel number of the top/bottom column. Using (7), the expression of the cost (5) becomes:

$$M(i,u) = M(i,u) - log(P(i,u)) \quad (8)$$

When a low number of interest points are matched by the detector, this amendment is not strong enough to guide the optimal path through the large occlusion regions. We have added a secondary cost, which penalizes the paths that don't follow the direction to the pivots.

Given (x_s, y_s) a matched pair and (u, i) the current position, the extra cost for each possible direction, which will be added to (8) is given by:

$$\begin{cases} C1 = C(i+1,u) = atan(\frac{x_s - i}{y_s - u}) \\ C2 = C(i,u+1) = atan(\frac{y_s - u}{x_s - i}) \\ C3 = C(i+1,u+1) = \left|45 - C1\right| \end{cases} \quad (9)$$

The influence of the additional cost for guiding the optimal path can be observed in Figure 8. The tests showed improved results in large occlusion regions.

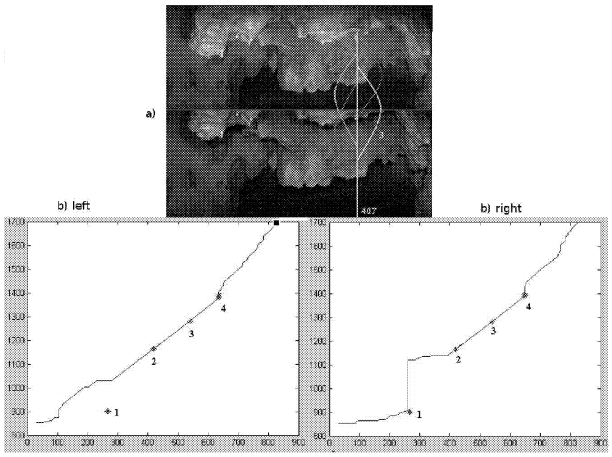


Figure 8: *Optimal path found using the guided Viterbi for the* 407^{th} *column, highlighted in a). It contains an important occlusion region (150 pixels). The coloured lines in a) indicate the interest points matched by the detector, marked in red in b). In b) left, the path can't follow the pivots. The improvement brought by the guiding cost is observed in b) right.*

4 CONCLUSIONS AND FUTURE WORK

We have presented an original catadioptric stereoscopic system, which uses a single linear CCD and a mirror system to obtain two projections of the scene in the same image, one above the other. An image issued by PCam-ST contains the necessary information for achieving a 3D reconstruction of the scene. The dedicated image matching algorithms use a 1D approach. The image matching was carried out firstly for detecting and matching interest points, and then it was extended to a guided dense matching, using the dynamic programming. The results were validated using noisy images, with large occlusion regions. The Viterbi algorithm is suitable for implementing a 1D dense matching, proving robustness to noise. As for the large occlusions, the guiding proved its efficiency. Improvements should be brought on the accuracy of the 1D interest point detector. Also, the results could be ameliorated by implementing a 2D smoothing over the Viterbi solution.

AKNOWLEDGMENTS

All the mechanical and optical design of the PCam-ST camera has been carried out by Philippe Puech, engineer in our team. He defined also a dedicated calibration procedure for this type of camera. Many thanks, Phil!

REFERENCES

Brown M.; Burschka D.; and Hager G.D., 2003. *Advances in Computational Stereo.* In *Transactions on Pattern Analysis and Machine Intelligence.* IEEE, 993–1008.

Cox I.; Hingorani S.; and Rao S., 1996. *A maximum likelihood stereo algorithm.* In *Computer Vision and Image Understanding.* 542–547.

Criminisi A.; Shotton J.; Blake A.; Rother C.; and Torr P.H.S., 2003. *Efficient dense stereo and novel-view synthesis for gaze manipulation in one-to-one teleconferencing.* In *Technical Report MSR-TR-2003-59.* 41.

Geiger D.; Ladendorf B.; and Yuille A., 1995. *Occlusions and binocular stereo.* In *International Journal of Computer Vision.* 211–226.

Harris C. and Stephens M., 1988. *A combined corner and edge detector.* In *Proceedings of the fourth Alvey Vision Conference.* 147–151.

Lowe D., 2004. *Distinctive Image Features from Scale-Invariant Keypoints.* In *International Journal of Computer Vision.* 91–110.

Nayar S., 1988. *Sphereo: Recovering depth using a single camera and two specular spheres.* In *1988 Cambridge Symposium on Advances in Intelligent Robotics Systems.*

Nayar S. and Baker S., 1998. *A theory on catadioptric image formation.* In *1998 Sixth International Conference on Computer Vision.* 35–42.

Southwell; Basu A.; Fiala M.; and Reyda J., 1996. *Panoramic stereo.* In *ICPR '96: Proceedings of the 1996 International Conference on Pattern Recognition.* 378.

Torr P. and Criminisi A., 2004. *Dense stereo using pivoted dynamic programming.* In *Image and Vision Computing.* 795–806.

Viterbi A.J., 1967. *Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. IEEE Transactions on Information Theory,* 260–269.

Generating Full Length Impaired Movies for Quality of Experience Assessments

Sebastiaan Van Leuven, Glenn Van Wallendael, Peter Lambert, and Rik Van de Walle
Ghent University - IBBT
Department of Electronics and Information Systems - Multimedia Lab
Gaston Crommenlaan 8 bus 201, B-9050 Ledeberg-Ghent, Belgium
email: {sebastiaan.vanleuven, glenn.vanwallendael, peter.lambert, rik.vandewalle}@ugent.be

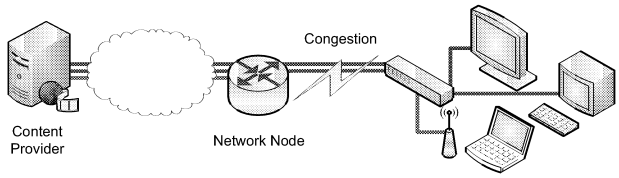
KEYWORDS
Quality of Experience, Scalable Video Coding

ABSTRACT

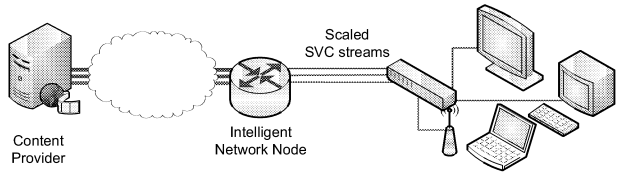
One of the latest advances in video coding is the scalable extension of the H.264/AVC video coding standard (SVC). This new development creates a lot of opportunities for content providers to adapt video streams to the needs of the end user, to the capabilities of the end users' playback device, and to the conditions of the intermediate transport network. Consequently, these adaptations will influence the characteristics of the decoded video. To investigate how these adaptations can be done best, research on the impact on the quality of experience (QoE) has to be conducted.
For the assessment of full length movies, or to create impaired streams, currently no framework exists. We propose a methodology to generate full length impaired movies for QoE research using SVC.

INTRODUCTION

Subjective video quality assessments are an important research topic to evaluate the viewers' appreciation of the quality of the video. These tests are mostly performed on short sequences, which viewers have to rate according to the percepted quality, and are mainly conducted in a protected environment with standardised equipment, such that the conditions between different viewers are the same. In these tests, users are very focused and are expecting visual errors, therefore these tests might not be completely accurate to investigate the quality of experience (QoE) in real life situations. The QoE is how the user experiences the perception of media. This QoE highly depends on his personal habits and how the user is used to percept the video.
Previous work (Staelens et al. 2009) has shown that viewers notice fewer artifacts in an impaired sequence when they are not mainly focused on quality assessment, and it has shown that for the evaluation of the influences on the QoE, it is preferable to use full length movies. If the user would be able to conduct these tests in his normal environment, the influence on the actual perception can be measured. This way, viewers might not notice small visual errors because they are not concentrating on the possible distortion of the video.
With the advent of the recently standardised scalable video coding (SVC) extension for H.264/AVC (Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 2007a), new possibilities to handle network problems arise. During transport, a scalable video stream can be adapted to the varying network conditions. This implies a degraded QoE but the exact impact depends on how scalability is included during encoding and how the original stream is altered.
It is important to notice that scalability is not only designed for adaptation due to varying network conditions, also adaptations depending on the needs of the end-user were kept in mind during standardisation. It is clear that a user with a high-definition television needs a different stream compared to a user who is using a mobile phone to watch the same video content. However, the scalabilities designed to support these situations are also exploited for adaptations due to network congestion.
Figure 1(a) depicts a situation where a network is congested. In this case, three different streams should be delivered, for which transmitting the complete SVC streams results in a congested network and subsequent uncontrolled packet drop. To solve this problem, an intelligent network node can scale the requested SVC streams, as can be seen in Figure 1(b). This way, every user receives a part of the requested content, moreover, no random packets are dropped, but the network node rejects the packets that will have the least impact on the QoE. To identify these packets, research still has to be conducted. Out of that research, a model for an intelligent node could be presented.
As far as the authors' knowledge goes, no framework for full length video quality assessments exists. To satisfy this need, we propose a methodology to create impaired streams for QoE research, that can be performed at the subjects' home. A full length movie is encoded using SVC, thereafter the effects of network conditions are simulated. The impaired SVC stream is converted to a DVD movie. These discs are distributed to viewers who can watch them in their normal environment, and rate them according to the perceived quality.
This methodology can be used for research about stream shaping, due to an intelligent node as shown in Fig-



(a) Single-layer stream dropped.



(b) Intelligent node scales SVC stream.

Figure 1: SVC Stream Adaptation Inside the Network.

ure 1(b), research concerning error concealment methods, or any other research using full length movies.

In the next section we will go into detail about SVC, so that the implications of scaling are clarified, thereafter the methodology is presented and possible modifications are proposed. Finally, a conclusion is given accompanied with future work that can be carried out.

SCALABLE VIDEO CODING

The recently developed and finalised SVC standard allows an encoded video stream to be scaled (Schwarz et al. 2007). This scaling means that a sub-stream, which is still decodable, can be extracted. This yields a stream with a reduced bit rate, resolution, or frame rate. Doing so, less bandwidth is required to transmit the sequence, less memory and computational power is needed to decode the sequence.

The concept of layers is introduced to denote the used scalability. The sub-stream containing the lowest resolution and least quality is referred to as the base layer. To extend the information of the base layer, one or more enhancement layers are added. A layer extending the information of the current layer is referred to as a higher layer.

Prior to encoding, the scalability has to be specified. Afterwards, only the scalabilities included in the encoded stream can be exploited by the network or application. The three main types of scalability that are supported in SVC are spatial, temporal, and quality scalability; these types can also be combined. Spatial scalability is achieved when a sub-stream represents the content at a lower spatial resolution. Temporal scalability provides a sub-stream with a lower frame rate (i.e., temporal resolution). Quality scalability allows the extraction of a sub-stream, for which the decoded content is reduced in picture quality (i.e., a reduced PSNR).

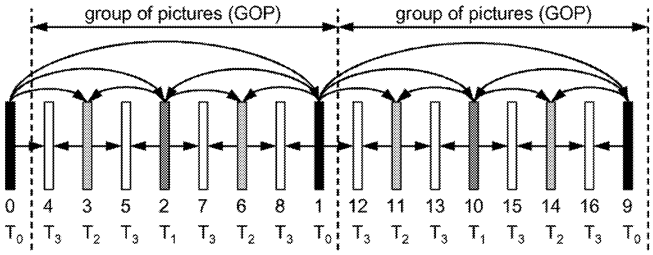


Figure 2: Example of a hierarchical prediction structure.

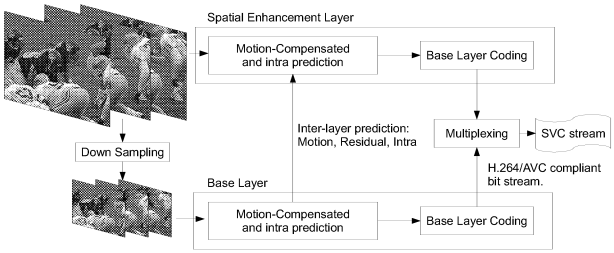


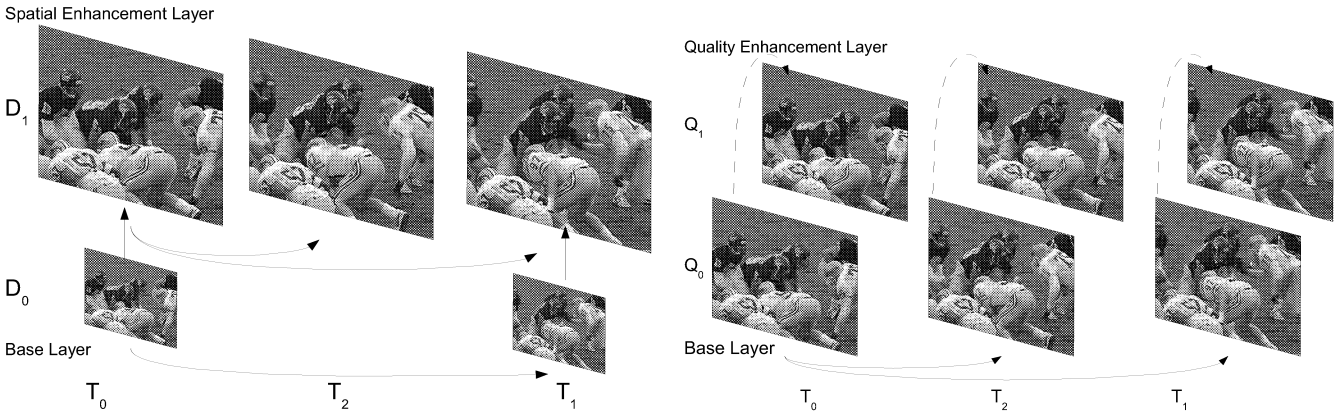
Figure 3: SVC Coding Structure.

Temporal Scalability

Temporal scalability allows to extract a sub-stream with a reduced frame rate. To achieve this, each frame is assigned to a temporal layer, identified by (T). In order to come up with a stream where frames are skipped at regular intervals, a hierarchical prediction structure inside a group of pictures (GOP) is used for encoding, as shown in Figure 2. The prediction only depends on frames of lower temporal layers. This implies that the prediction is not necessarily done on the previous frame in display order. Rejecting frames of higher temporal layers result in a sub-stream with a reduced frame rate.

Spatial Scalability

Adding a spatial enhancement layer (identified by D) increases the resolution of the content compared to the previous layer. It is not permitted that a spatial enhancement layer has a reduced spatial resolution in comparison with a lower one.
Spatial enhancement layers can be encoded with completely different characteristics compared to other spatial enhancement layers, as an illustration two spatial layers might use a different entropy encoding. Each spatial layer is treated as it would be encoded as a single layer, so that for each layer intra and motion-compensated prediction is performed, as is visualised in Figure 3. In order not to end up with a simulcast scenario (where each spatial layer is treated as a separate video stream that is encoded independently and transmitted simultaneously), visual correlations between the current spatial enhancement layer and its dependency layer can be exploited using inter-layer prediction mech-



(a) Temporal and Spatial Scalability

(b) Temporal and Quality Scalability

Figure 4: Combined Scalability with SVC.

anisms. Three types of inter-layer prediction are allowed in SVC: inter-layer motion prediction, inter-layer residual prediction, and inter-layer intra prediction (Segall and Sullivan 2007).

Quality Scalability

Coarse Grain Scalability
Quality scalability can be obtained by using spatial layers with the same resolution and assign each spatial layer a different quantisation parameter (referred to as coarse grain scalability, CGS). However, using CGS yields to a low flexibility, because only at fixed points in the stream it is possible to switch to higher spatial enhancement layers Furthermore, a higher bit-rate is obtained because of the syntax and encoding overhead that is introduced due to the inter-layer prediction.

Medium Grain Scalability
Medium Grain Scalability (MGS) uses quality enhancement layers, identified by Q. In total 16 quality layers can be used for each spatial enhancement layer. This results in a flexible stream where, for each frame, the quality can be adapted by reducing the number of quality layers. Doing so, less residual information will be available for decoding, such that the decoded picture results in a lower visual quality. No predictions are used between quality layers, therefore the decoding overhead is limited.

Figure 4 gives an example of combined scalability, using the above mentioned types of scalabilities. The continuous lines indicate the used predictions, the dashed lines indicate the quality enhancement layer. In Figure 4(a), a combination of temporal and spatial scalability is shown. The base layer has half the frame rate compared to the spatial enhancement layer. When T_2 is dropped, the enhancement layer is still decodable because of the hierarchical prediction. The MGS in Figure 4(b) is not using inter-layer prediction. Furthermore, it is possible to adapt the quality for each frame, because no predictions rely on frames where $Q = 1$.
As with H.264/AVC, an encoded SVC stream is packetised to form Network Abstraction Layer units (NAL units). In the header of each NAL unit, the layer identifiers ((D, T, Q) triplet) are transmitted. Next to this triplet, also a *discardable_flag* is transmitted which indicates if the current NAL unit is used for the decoding of frames in a higher spatial layer. These syntax elements will later be used in our methodology.

METHODOLOGY

In this section, we describe how a system for generating full length impaired movies needs to be designed. In particular a methodology is presented. Figure 5 gives a schematic overview of this methodology,keeping in mind the use case of QoE assessments.
The full length quality assessments will be conducted at the subjects' home. Therefore, the most preferable output format is a DVD disc. This disc has to be playable with any conventional DVD player, this way the normal television watching conditions are closely matched. The resulting disc contains an impaired version of an original movie. How the original movie is altered and how the disc is created is highlighted below.

Overview

The input for the impaired movie is a DVD disc containing a full length movie. Because the information on a DVD is encrypted and placed inside .vob files, first these files have to be copied from the disc and decrypted. Later on, the movie itself is extracted from the .vob files, such that it can be encoded to an SVC stream. The encoded stream is tampered and decoded. The decoded result is converted to a DVD compatible format. At last, the disc is written. Details on the error introduc-

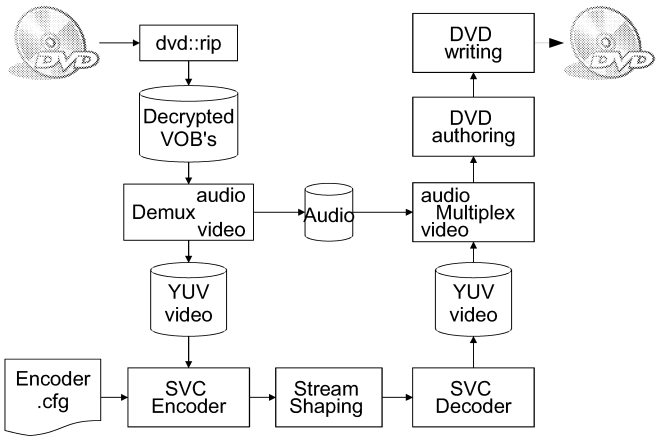


Figure 5: Schematic Overview of the Methodology.

ing step are visualised in Figure 6. In the remainder of this section, we will go into detail in the steps that have to be performed.

Extraction

The extraction of the encrypted .vob files can be done using freely available software (dvd::rip Software 2008) or (vobcopy Software 2008). The resulting decrypted .vob files contain an MPEG Program Stream, which includes the (MPEG-2) video stream and audio stream. Because audio is not handled within SVC, the next step is to demultiplex the audio from the video, such that the video can be used for encoding with SVC. Demultiplexing is done using (Mplayer Software 2009). Because the audio file is not modified, the resulting movie has synchronised audio and video, as long as the number of frames of the original stream equals the number of frames of the impaired stream.

Encoding

For encoding a video using SVC, reference software is available, which is referred to as Joint Scalable Video Model (JSVM) (Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 2007b). This software is used for encoding the movie to an SVC compliant bit stream. Before continuing, a few considerations have to be made. First, it is recommended to use a software version prior to JSVM version 9.10, because from version 9.10 on, error concealment techniques in the decoder are not supported anymore. Since errors are introduced in the resulting bit stream, it is recommended to use an error concealment technique at the decoding side. Second, the input for the encoder is expected to be an uncompressed YUV-sequence (using 4:2:0 subsampling). To obtain this, the video inside the decrypted .vob files is decoded to a YUV sequence using (FFmpeg Software 2009).

We encoded the YUV sequence using JSVM version 9.1. The input for this process is an encoder configuration file, which gives the encoder the necessary information. As depicted in the JSVM-software manual (Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 2007b), the encoder configuration file should include, next to global stream configuration parameters, for each spatial layer a separate layer configuration file with parameters. These files describe the properties and the scalabilities that later could be exploited for these spatial layers. The resulting stream is a packetised SVC encoded stream.

Stream Shaping

A tool, that is able to parse the encoded stream, could be developed to adopt the stream in a way that real-life packet drops are simulated. However, a more flexible solution exists by using the 'bitstream extractor' tool, included in the JSVM reference software. The dual pass mechanism of this tool is visualised in Figure 6. Using the print trace option of the bitstream extractor (first pass), a trace file is generated which includes information for each NAL unit on each line, cfr. 'Original Trace File' in Figure 6. For each NAL unit the start position, the length, (D, T, Q)-triplet, packet type, and the remark if the NAL unit is discardable and truncatable are given. In a second pass, this tool allows to extract a sub-stream using a modified trace file as input, cfr. 'Impaired Trace File' in Figure 6. To achieve this, the bitstream extractor copies only those packets described in the modified trace file in the resulting bit stream based on the given start position of the packets in the original stream.
The information that is removed from the modified trace file corresponds to the packets that are discarded in the newly generated stream. This way errors are introduced in an artificial way. The modified trace file facilitates the way errors are simulated since an abstraction is made between the packet drop and the stream.
Different kind of errors can be introduced in the packetised stream, depending on the research topic. For a stream scaling situation, such as in Figure 1 (b), lines corresponding to NAL units of higher layers may be dropped. For error concealment research, a random packet drop of non-discardable NAL units can be done by removing random lines.
Despite the fact that any random line of information can be removed from the trace file, it is not recommended to do so. The resulting sub-stream has to be decodable, therefore stream headers and parameter sets have to remain in the modified trace file. Furthermore, only discardable NAL units should be removed. Depending on the research that has to be performed, different approaches can be used to remove information. When an intelligent network component, that removes certain layers, is simulated, NAL units should be removed depend-

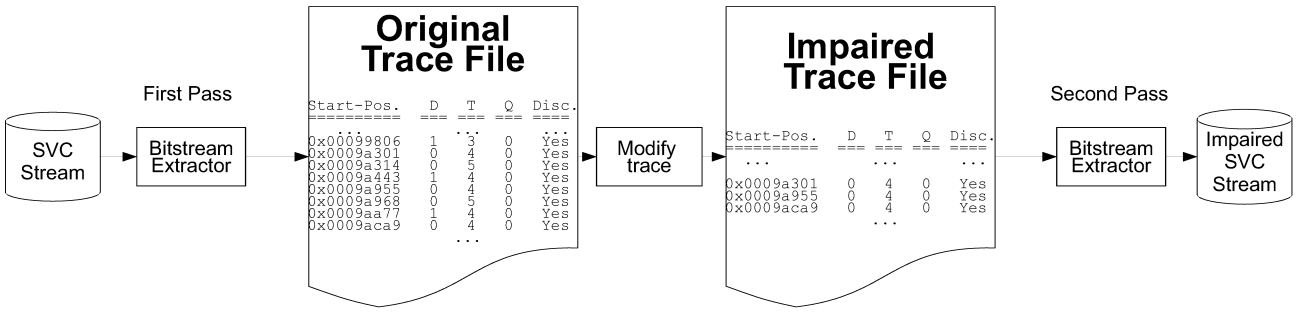


Figure 6: Trace File Modifications.

ing on the specified (D, T, Q) parameters. Only lines with higher layer identifiers are removed. For the evaluation of error concealment techniques, a possible strategy is to remove random lines from the original trace file. Anyhow, using the bistream extractor tool, developers have the freedom to implement removal schemes depending on their research. In order to allow inter-layer prediction mechanisms to work, packets of dependency layers used for prediction of available higher layers should not be removed. These packets might only be removed if the higher layers are also removed.
To mimmic the network conditions, the trace file is altered, according to above considerations. In Figure 6 this is shown in the Modify trace step, resulting in the Impaired Trace File. It can be seen that for the impaired sequence only the lowest resolution layer is allowed. Additionally, a reduction in frame rate has been achieved by removing the packets where $T > 4$. As can be seen, the impaired trace file only contains packets for which: $D = 0$ and $T \leq 4$.
Using this impaired trace file, the bitstream extractor generates an impaired SVC stream. From the point of view of a network node, this is an extracted sub-stream.

Decoding

The extracted sub-stream is decoded using the JSVM reference software decoder, which contains error concealment methods. The error concealment is used to guarantee that the total number of frames of the input and output sequences are the same, so that the audio and video are synchronised. The decoder outputs a YUV sequence.

Authoring and Writing

To end up with a DVD compatible file, the YUV sequence is encoded to an MPEG-2 stream, meanwhile the dumped audio track is copied in the MPEG Program Stream. This multiplexing is done using FFmpeg. Before writing the resulting full length impaired movie to a DVD disc, the DVD has to be authored first. This is done to make sure that the disc can be played with any conventional DVD player. In this process the necessary DVD structure and files are generated. Afterwards, the folder containing the authored dvd structure is written to a DVD disc. For these steps, many software tools are available; such as (dvdauthor Software 2007) for authoring and (growisofs Software 2008) for writing.

POSSIBLE MODIFICATIONS

The methodology described above is a general approach, nevertheless, a few modifications can be done towards optimisations of the required time and storage space.

One optimisation is to reduce the size of the YUV sequence. If the tests to conduct do not require the full stream to be adapted, but only a few impairments throughout the movie are desired, the movie does not have to be decoded completely. In this case, it is not necessary to decode all .vob files, but only those required to have impairments can be decoded. Furthermore, if only short sequences have to be available as YUV, an offset and number of frames can be specified, such that not the complete .vob has to be decoded. In this case special care has to be taken when multiplexing the audio and video streams. The impaired parts have to be inserted at the corresponding location in the stream. This can be specified with the command arguments using FFmpeg.

Additionally, when using an operating system supporting pipes, different intermediate files can become superfluous by using (named) pipes. When using pipes, it is possible to pass data between multiple processes without writing it first to a disc, and thus reducing the required storage space.

Another optimisation can be found in the time necessary to encode the sequence. Since the JSVM reference software is single threaded, the advantages of multi-core architectures cannot be exploited. This can be improved by splitting the YUV in multiple parts, and encode each part simultaneously on a different processor. Finally, also the decoder can take advantage of this approach so that also the decoding time is reduced.

To reduce the number of user interactions, the presented methodology can be automated using scripting languages.

IMPLEMENTATION

The proposed methodology has been implemented on a dual core computer running CentOS (Linux), using the above mentioned software. A full DVD movie (pal, 118 min.) has been decrypted, stored and demultiplexed. To reduce the storage space and encoding time, only the first three minutes have been converted to YUV and encoded. Quality scalability was used to illustrate the stream shaping capabilities.

After decoding, the resulting YUV sequence has been converted to an MPEG program stream and concatenated with the rest of the movie. Meanwhile the audio track was multiplexed to form an MPEG transport stream. Finally, this MPEG transport stream was authored and written to disc.

The implementation was automated using shell scripting, as this was possible by the operating system environment. Nevertheless, to extract the movie from the disc using dvd::rip, the correct title first has to be known, which is discovered using vobcopy. Once this is found, the correct title number is specified as an argument for the vobcopy command.

No time measurements have been made, because these would not have had any significance to the methodology. Since the JSVM encoder is not optimised for speed, the most time consuming task is undoubtedly the encoding step.

CONCLUSION AND FUTURE WORK

Using our proposed methodology, impaired SVC streams, that can be used for QoE research with full length movies, can be generated. The described methodology can be fully implemented using freely available software. Additional improvements are proposed, such that the required encoding time and storage space is reduced.

The methodology is not only limited to a scenario for research about the best way of altering SVC streams, also other QoE assessments can take advantage of this methodology, for example error concealment research.

In the future, a tool can be developed to modify the trace files, which will improve and facilitate the simulation of an intelligent network component.

ACKNOWLEDGMENTS

The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

REFERENCES

dvdauthor Software, 2007. `http://dvdauthor.sourceforge.net`.

dvd::rip Software, 2008. `http://www.exit1.org/dvdrip`.

FFmpeg Software, 2009. `http://www.ffmpeg.org`.

growisofs Software, 2008. `http://fy.chalmers.se/~appro/linux/DVD+RW/tools`.

Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2007a. *Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), Version 8 (including SVC extension).* Tech. rep., MPEG / ITU-T.

Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2007b. *Doc. JVT-Y203: Joint Scalable Video Model 12* . Tech. rep., MPEG / ITU-T.

Mplayer Software, 2009. `http://www.mplayerhq.hu`.

Schwarz H.; Marpe D.; and Wiegand T., 2007. *Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Transactions On Circuits and Systems for Video Technology*, 17, no. 9, 1103–1120.

Segall C. and Sullivan G., 2007. *Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension. IEEE Transactions On Circuits and Systems for Video Technology*, 17, no. 9, 1121–1135.

Staelens N.; Vermeulen B.; Moens S.; Macq J.F.; Lambert P.; Van de Walle R.; and Demeester P., 2009. *Assessing the Influence of Packet Loss and Frame Freezes on the Perceptual Quality of Full Length Movies.* In *International Workshop on Video Processing and Quality Metrics for Consumer Electronics.*

vobcopy Software, 2008. `http://vobcopy.org`.

AUTHOR BIOGRAPHY

SEBASTIAAN VAN LEUVEN received the M.Sc. degree in Applied Engineering from the University College of Antwerp, Antwerp, Belgium, in 2006 and the M.Sc. degree in Computer Science Engineering from Ghent University, Ghent, Belgium in 2008. In 2008 he was awarded for his thesis with the Barco Award. Currently, he is with Multimedia Lab, Ghent University, where he is working towards a Ph.D., with the financial support of the Interdisciplinary Institute for Broadband Technology (IBBT). His main research topic is video coding, including scalable video coding.

GLENN VAN WALLENDAEL received his Masters degree in Applied Engineering from the University College of Antwerp, Belgium, in 2006 and his Masters degree in Computer Science from Ghent University, Belgium in 2008. Since then, he has worked as a Ph.D. researcher at Multimedia Lab, Ghent University, and IBBT. His research interests are in the areas of multimedia signal processing and video coding.

PETER LAMBERT is a senior researcher at the Multimedia Lab of Ghent University - IBBT. He received his Master's degree in science (mathematics) and in applied informatics from Ghent University in 2001 and 2002, respectively. He obtained the Ph.D. degree in computer science in 2007 at the same university. His research interests include multimedia applications, (scalable) video coding technologies, and error robustness of digital video.

RIK VAN DE WALLE received the M.Sc. and Ph.D. degrees in engineering from Ghent University, Ghent, Belgium, in 1994 and 1998, respectively. After a visiting scholarship at the University of Arizona, Tucson, he returned to Ghent University, where he became a Professor of multimedia systems and applications, and Head of the Multimedia Lab. The Multimedia Lab is also one of the partners of the Interdisciplinary Institute for Broadband Technology (IBBT), which was founded by the Flemish Government in 2004. His current research interests include multimedia content delivery, presentation and archiving, coding and description of multimedia data, content adaptation, interactive (mobile) multimedia applications, and interactive digital TV.

THE NEW DELFT UNIVERSITY OF TECHNOLOGY DATA CORPUS FOR AUDIO-VISUAL SPEECH RECOGNITION

Alin G. Chițu[1] *and Leon J.M. Rothkrantz*[2]

[1, 2]Man-Machine Interaction Group, Department of Mediamatica
Delft University of Technology, Delft, The Netherlands
[1]a.g.chitu@tudelft.nl, [2]l.j.m.rothkrantz@tudelft.nl

KEYWORDS

Data Corpus, DUTAVSC, NEWDUTAVSC, Speech Recognition, Lip reading, Active Appearance Models.

ABSTRACT

In a world where we face a continuous spread of technology and communications but also a world of catastrophic events and terrorism which leads to the need of increased security, the lip reading technology gets increasing attention in the scientific community. This paper details on our work towards a robust automatic lip reading system. We proudly introduce the data corpus for Dutch language we have built for the present research with details on its development and characteristics. This paper also introduces the use of Active Appearance Models for the detection of landmarks on the speaker face which we are going to use for lip reading.

INTRODUCTION

As speech recognition came to a maximum of performance and good recipes for building speech recognizers were written, the need to search for complementary information from other sources came naturally. Lip reading can thus be seen as a complementary process to speech recognition but also as stand alone process. The trend is to make the communication between humans and their artificial partners easier and more human like (e.g. Ambient Intelligence (AmI)). The applications for lip reading as stand alone applications are multiple: multimedia phones for the hearing impaired, mobile phone interface for public spaces (e.g. at the time of this writing, phone models that use lip reading are already being designed,) person identification, recovery of speech from deteriorated or mute movie clips, and, maybe the most promoted applications, security applications (e.g. security cameras that are recovering what is being said without the need of microphones and more important the possibility to understand what is being said from large distances.)

In this paper we introduce in detail our data corpus, namely recording settings and corpus characteristics. However, we are also providing insights on the feature extraction methods we are going to use for processing the data. We split our paper in three obvious parts data acquisition which gets the most space, data parameterization and inference engine. Hence data acquisition is presented in great detail in Section DATA ACQUISITION. The importance of the data corpus gets an entirely new scale when one starts building a data corpus. The amount of work involved is astonishing, which entitles the process for better visualization (we spent almost an entire year.) However, the result was the largest and most complex bimodal data corpus for Dutch language, to date. The second aspect is data parameterization. The parameterization of the input data is twofold; it should reduce the dimensionality and capture the relevant information about the process being modeled. Therefore, there are many techniques being employed here. In our current work we chose to use landmark points which are computed based on Active Appearance Models [1] (AAM.) Section ACTIVE APPEARANCE MODELS introduces the AAM methodology and the landmarks we have chosen with their definitions. The visual features are introduced in Section VISUAL FEATURES. For inference about what is being said we used Hidden Markov Models approach and the HTK Toolkit [2] developed at the Speech Vision and Robotics Group of the Cambridge University Engineering Department. Details about building the recognition models are found in Section INFERENCE ENGINE. We conclude this paper by describing the remaining work to be done to finalize our efforts in Section ONGOING WORK AND FUTURE EXPECTATIONS.

RELATED WORK

Lip reading literature has increased rapidly over the years. There are two possible directions of research namely, data parameterization and inference mechanisms.

Due to the temporal characteristics of lip reading the majority of techniques used for other machine learning areas are not suitable here as in the case of speech recognition. However, for limited vocabulary applications it could still be appropriate. Wang et al. in [3] used b-spline functions to match the spoken utterances. Other similar approaches used PCA analysis. For large vocabulary continuous speech lip reading as in the case of speech recognition a time series approach

is suitable. Also due to its success in speech recognition, by far the most used method is Hidden Markov Models with its derivates.

On the other side for feature extraction there were many methods developed. They fit mainly in two broad classes: appearance based methods and geometrical methods; a combination of the two was also used. The approaches from the first class consider the raw image or a transformation of it as feature processing [4, 5]. The transformation of the image is employed in order to obtain some data reduction. The most popular method for this is Principal Component Analysis (PCA) [6, 7]. Other methods which were used as an alternative to PCA are based on discrete cosine transform [8] and discrete wavelet transform. However, this approach gives rise to very high dimensionality of the feature vectors. On the other side the algorithms try to model the speech production apparatus. However, not all parts of the speech production system are visible, hence these methods try to model the visual parts of the speech apparatus such as lips, teeth, tongue but also other parts of the face. The algorithms are aiming to detection and tracking of specific points on the face. Usually the detection process is assisted by 2D or 3D geometrical models of the face [9]. One other approach is to use statistical methods based on image filtering algorithms and try to directly extract the shape of the lips. The dimensionality reduction obtained through the latter approaches is very large. Moreover the reduction of dimensionality was done in a direction that is more appropriate for speech recognition. Another approach is to perform optical flow analysis on the input video stream both as a measure of the overall movement on the face and used only for onset/offset detection [10, 11, 12, 13], or as feature generation engine as in [14].

DATA ACQUISITION

Data corpus is the building block of any successful research. After working for some time with our old small data corpus we arrived to the conclusion that a new larger and stronger data corpus is needed. We did extensive analyses to discover what the drawbacks of the existing corpora were. The paper [15] provides a rich comparison among some of the most used data corpora and introduces a set of guidelines to be followed when building a data corpus. Some of the important issues spotted were the resolution of the video recordings, the frame rate of the video recordings, the richness of the language pool and not the least the size of the corpus. The paper [16] analyzes the issue of the video frame rate influence on the information retrieval. The conclusion drawn by the authors was that in order to accurately capture all the information emanating from the speaker face we need to record at higher frame rates than what standard consumer cameras currently provide. Figure 1 shows the poor coverage of the visemes in the case of fast speech rate. The results are based on the Delft University of Technology Audio-Visual Data Corpus (DUTAVSC) [17], which we previously used in our research. Hence, in the case of fast speech rate the data becomes very scarce; we have a mean of 3 frames per viseme which can not be sufficient. Therefore, building a new data corpus became the new point on our agenda. The resulted corpus is to our knowledge the largest and most complete bimodal corpus for Dutch language to date. The data corpus will be made available to the scientific community in the near future.

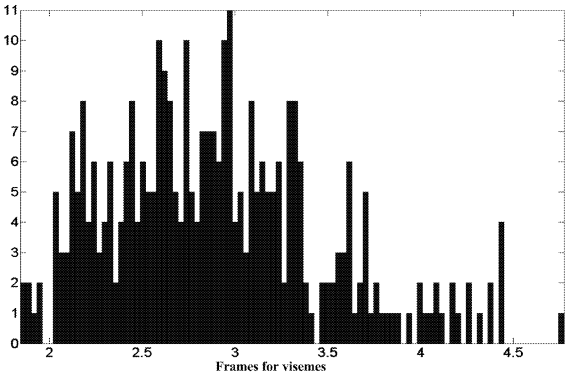


Figure 1. Viseme coverage in the case of fast speech rate.

Recording settings

We considered recording in controlled environment (i.e. reasonable noise levels and good illumination.) The specific noise can be simulated or recorded in the required conditions and later superimposed on the clear audio data. An example of such database is NOISEX-92 [18]. This dataset contains white noise, pink noise, speech babble, factory noise, car interior noise, etc. On the video side we used a uniform background so that by employing chroma keying techniques one could immerse the speaker in any environment (e.g. weather forecast on television.)

One more important improvement is adding a side view to the recordings. We devised two settings to achieve this effect. Firstly we used a 45 degree placed mirror; we eventually gave up this idea due to image distortions. We settled in the end to a second idea which used 2 cameras to capture both frontal and side view. We used two very sensitive condenser microphones one for recording the speaker voice and a second to record the background noise. The audio signal was sampled at 48 kHz and used a sample size of 16 bits. The video was recorded at 100Hz for both side and frontal view, half PAL resolution. Only the lower face was recorded.

Using a high speed camera increases the storage needs for the recordings. It is almost impossible to record everything and than in the annotation post

process cut the clips at the required lengths. One main reason is that when recording in high speed - high resolution the bandwidth limitation requires that the video be captured in the memory (i.e. RAM Drive) or on a RAID0 stripe system. The first approach limits the length to approximately 1 minute depending on the resolution and color subsampling ratio used. The second approach removes the time limitation but is not always reliable, due to the operating system. However, we anyway needed to present the speakers with the pool of items required to be uttered. We build therefore a prompter like tool that provided the user the next item to be uttered together with some instructions about the speaking style and also controlled the video and audio devices. The result was synchronized audio and video clips already clipped to the exact length of the utterance. The tool provided the speaker the possibility to change the visual themes to maximize the visibility, and offer a better recording experience. Figure 2 shows a screenshot with the tool.

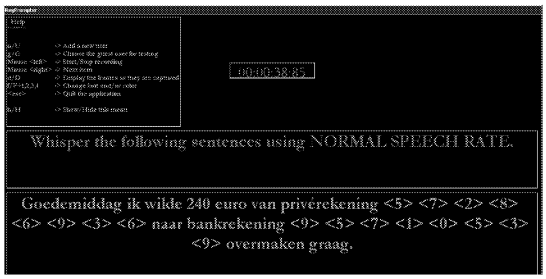


Figure 2. Prompter which instructed the speaker about the required speaking style and the next utterance.

Statistics about the data corpus

The language pool contains 1966 unique words, 427 phonetically rich unique sentences, 91 context aware sentences (i.e. banking applications), 72 conversation starters and endings and 41 simple open questions (i.e. for these questions the user was asked to utter the first answer that they think of. In this way we expect to collect more spontaneous aspects of the speech.) For each session the speaker was asked to utter 64 different items (sentences, connected digits combination, random words, and free answer questions) divided in 16 categories with respect to the language content and speech style: normal rate, fast rate and whisper.

The corpus resulted needs a storage memory of 473GB. We recorded 67 speakers: 20 female and 47 male. Three speakers recorded 2 sessions, therefore a total of 70 sessions were recorded. Each session has in average 60+ utterances recorded. The total number of utterances is above 4200, amounting for more the 6 hours of continuous recording.

For each speaker we recorded the gender, age, education level, native language (as well as whether he/she is bi-lingual) and region where he/she had grown up. The last aspect is used to identify possible particular clusters determined by the regional dialects. In the paper [19] is shown that the contextual background of a speaker influences both the content and the aspect of the speech. The recognition of the context could enable us to build adaptive systems which can provide increased accuracy and reliability. The speakers were in large majority native Dutch speakers; hence we expect that the cultural impact is reduced.

ACTIVE APPEARANCE MODELS

Active Appearance Models (AAM) technology is employed to extract the location of specific points on the face from every frame of the video sequence. AAM was introduced in the paper [20] and is a generalization of the Active Shape Models (ASM) and combines both shape information (i.e. model based approach) and texture information (i.e. appearance based approach) in its searching scheme.

AAM creates a statistical model of shape variation and a model of texture variation. The average shape is determined considering the training set of shape samples. The shape samples are aligned using a Generalized Procrustes Analysis. Each face sample is then warped so that the control points match the ones of the mean shape (Figure 3).

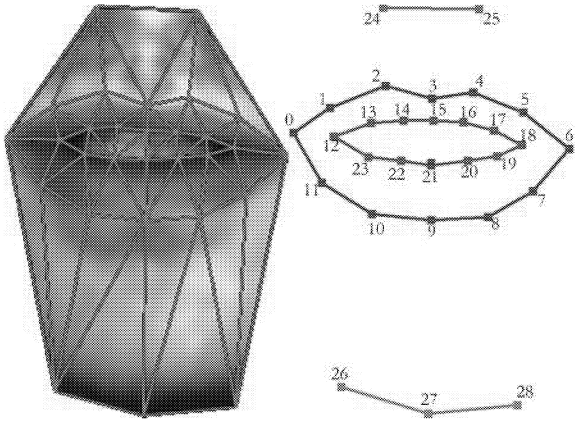


Figure 3. The mean shape and the mean texture used for AAM.

The search starts from the mean model and iteratively modifies the model parameters inside the learned range, while minimizing the difference in appearance between the real image and the image synthesized based on the new model (*Figure 4*). The required number of parameters is computed in both cases by using PCA. As with the majority of all searching schemes a good initial guess helps to speed

up the process. The use of a face/mouth detection/tracking algorithm as a prior step was found to greatly speed up the search for the shape parameters during AAM based processing. This enhancement made possible a real-time implementation of the algorithm.

Figure 4. Some search sequence, the first five and the optimal found after 34 iterations.

Landmarks and their definitions

The shape information extracted by the AAM from a face image is used to compute a set of suitable parameters that describe the appearance of the facial features. Therefore the first step is the selection of the optimal key points, relevant from the point of view of speech related movements. Kobayashi and Hara defined a set of Facial Characteristic Points (FCPs) extensively used for facial expressions recognition. We created our model based on Kobayashi & Hara model, however only paying attention to the mouth's region. The model used is shown in Figure 3. The model consists of 29 points distributed around the mouth, nose and chin.

The next step is to transform the FCP-set to some parameters of an intermediate model. The parameterization has the advantage of providing the classifier with data that encode the most important aspects of visual speech. This also acts as a dimensionality reduction procedure since the dimension of the feature space is lower than the dimension of the image space. As it is described in Section 0 the feature parameters are computed as certain Euclidean distances and areas between key points. Kobayashi and Hara also introduced a set of symmetry constraints which enable the recognition system to recover information when the image is degraded due to illumination or occlusion or to enable the uniform annotation. Our model uses the following constrains:

- the points 3, 9, 15, 21 and 27 should be on the same line
- the line formed by the points 24 and 25 should be orthogonal on the line formed by the points 3 and 27.
- some pairs of points are formed. The points in a common pair should follow a similar definition. These pairs are: (1, 11), (2, 10), (4,8), (5,7) , (13, 23), (14,22), (16,20) and (17,19)

Landmarks definitions

In order to keep the annotation of the video frames uniform (i.e. since the amount of data to be annotated is enormous, some more than 2 millions of frames, we divided the data among multiple annotators. Each annotator was asked to train and supervise the AAM.)

Outer mouth contour

The points on the outer mouth contour are defined as follows:

- the point 0 is the left most point, still on the lips (i.e. left mouth corner.)
- the point 6 is the right most point, still on the lips (i.e. right mouth corner.)
- the points 2, 3 and 4 are placed in accordance with the philtrum (infra-nasal depression), namely, 2 and 4 at the foot of the philtral columns, respectively, and 3 in the place where the philtrum meets the upper lip.
- the points 8, 9 and 10 are corresponding to points 4, 3 and 2, respectively.

The points 1, 5, 7 and 11 are placed such that the lips area is covered as closely as possible. However, their positions are preferred to be at equal distances from their neighboring points. It should be noted that there is more person-dependent information on the outer mouth contour.

Inner mouth contour

In the case of the inner mouth contour the decision was that the stress should be placed on accurately describing the aperture of the mouth. A special case is for a closed mouth. The points on the inner contour are closely related with the ones on the outer contour and have similar definitions.

- the point 12 is the left most point in the cavity of the mouth but not on the lips. However, in the case of a closed mouth this is not possible to observe. In this case this point should be placed such that is best describing the mouth line, but always to the left of the points 13 and 23.
- the point 18 is the right most point in the cavity of the mouth but not on the lips. In similar way as for the point 12 in the case of a closed mouth this is not possible to observe. In this case this point should be placed such that is

best describing the mouth line, but always to the right of the points 17 and 19.
- the points 15 and 21 are corresponding to the points 3 and 9 and follow the philtrum. Moreover all 4 points should always lie on the same line.

The last 8 points form pairs as follows 13 and 23, 14 and 22, 16 and 20 and 17 and 19, and they have similar definitions with the corresponding points on the outer contour.

Nose

The nose points are only a delimitation of the nose and are going to be used as a reference for computing distances to the other points (e.g. the distance from the point 27 on the chin to the line formed by the point 24 and 25 is used for onset/offset detection.) The points are placed at the base of the nose.

Chin

We are interested here to track the tip of the chin marked by the point 27. The point 26 and 28 are only helping the detection of the point 27 and should be placed symmetrical with respect to the point 27 and describe as closely as possible the chin. The point 27 should be aligned to the points 3, 9, 15 and 21.

VISUAL FEATURES

As we indicated in Section 0 the next step after computing the landmark points is to extract an optimal parameterization with respect to lip reading. This parameterization should best describe the curvature of the mouth and capture the transformations the mouth goes trough during speech. From the literature we find that mouth width and mouth height are two measures that were extensively used. To these two we added five other measures: mouth area, mouth aperture height, mouth aperture width, mouth aperture area and nose to chin distance. The next subsection gives our definition to all these measures.

Visual feature definitions

To have a unified understanding we include here the definitions of the visual features as we used them.

- **Mouth height** is defined as the Euclidian distance between the points 3 and 9.
- **Mouth width** is defined as the Euclidian distance between the points 0 and 6.
- **Mouth area** is defined as the area inside of the outer contour.
- **Aperture height** is defined as the largest Euclidian distance between the pairs of the points (13, 23), (14, 22), (16, 20), (17, 19) and (15, 21).
- **Aperture width** is defined as the Euclidian distance between the first point, or coinciding pair of points, and the last point, or coinciding pair of points, on the inner contour. The points are ordered left to right.
- **Aperture area** is the area covered by the mouth aperture, namely the inner contour.
- **Nose to chin distance** is the distance between the point 27 to the nose line (i.e. the line formed by the points 24 and 25).

INFERENCE ENGINE

For the actual recognition we used the Cambridge University Hidden Markov Model Toolkit. The recognition units were chosen to be the visemes. The visemes are the phonemes' corresponding semantic unit of speech in the visual domain. For Dutch language, most researchers use a set of 40 phonemes. However, since there is no one-to-one mapping from the phonemes set to the visemes set researchers define the visemes by clustering together the phonemes which have similar visual outcome. The decision about the similarity of the outcome is based on the degree of confusion and distinction human subjects are attaining [21, 22]. We used a set of 16 visemes shown in Table 1. The phonemes where tied into groups as defined in the paper [23].

Table 1. Viseme set for Dutch language in SAMPA notation.

#	Viseme	Phoneme set
1	[F]	f v w
2	[S]	s z
3	[X]	S Z
4	[P]	p b m
5	[G]	g G k x n N r j h
6	[T]	t d
7	[L]	L
8	[I]	I e:
9	[E]	E E:
10	[A]	A
11	[@]	@
12	[i]	i
13	[O]	O Y y u 2: o: 9 9: O:
14	[a:]	a:
15	[H]	H
16	[E+]	E+

To the set of visemes we added two extra models for *silence* and *short pause*, thus a number of 16 models are used. Each viseme is modeled by Gaussian mixtures continuous density left-right HMM with 5-states, with

only three emitting states. The model is shown in Figure 5. Special attention is paid to the silence models since in the case of visual data even in the case when nothing is uttered the signal energy can be still very high.

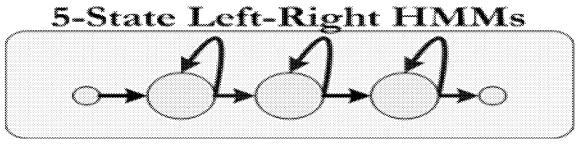


Figure 5. 5-State Left-Right HMM where only the middle states are emitting. The first and the last states are used for linking models together.

The models are trained iteratively, using embedded Baum-Welch re-estimation and Viterbi alignment.

ONGOING WORK AND FUTURE EXPECTATIONS

The processing of the data corpus proved to be very tedious since we dealt with a large number of frames, more than 2 million frames needed to be processed. Therefore, encouraged by the performance obtained with the AAM technology, namely the possibility to optimize the processing to almost automatic feature extraction, we started working with the geometric features introduced by this paper. Based on the result we obtained in the past with a much smaller data corpus, we expect to have better results. However, our belief is that other features should be added for even better result. As concluded in the paper [22] the presence of teeth, tongue is valuable information for lip readers. Therefore, their detection should be included [14]. More advanced features such as optical flow features should be investigated as well.

ACKNOWLEDGEMENTS

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024 file.

We thank all the participants to our long and tiring recoding sessions, but especially to Karin Driel, Pegah Takapoui and Mathijs van Vulpen for their valuable help in building up the language pool, setting the recording environment and supervising the recordings.

1. References

[1] Cootes, T.F., Edwards, G.J., Taylor, C.J., "Active Appearance Models", In *Proceedings of the 5th European Conference on Computer Vision* – Volume 2, 484-498, Springer, 1998.

[2] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X. A., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P., "The HTK Book (for HTK Version 3.4)", 2005.

[3] Wang, S. L., Lau, W. H., and Leung, S. H., "Automatic Lipreading with Limited Training Data.", In *Proceedings of the 18th international Conference on Pattern Recognition* Vol. 3, Washington, DC, 881-884, 2006.

[4] Li, N., Dettmer, S., and Shah, M., "Lipreading using eigen sequences", In Proc. International Workshop on Automatic Face- and Gesture-Recognition, (Zurich, Switzerland), pp. 30–34, 1995.

[5] Li, N., Dettmer, S., and Shah, M., "Visually recognizing speech using eigen sequences", *Motion-based recognition*, 1997.

[6] Hong, X., Yao, H., Wan, Y., and Chen, R.,"A PCA Based Visual DCT Feature Extraction Method for Lip-Reading", iih-msp, vol. 0, pp. 321–326, 2006.

[7] Bregler, C., and Konig, Y., ""Eigenlips" for robust speech recognition", in Acoustics, Speech, and Signal Processing, ICASSP-94 IEEE International Conference on, 1994.

[8] Duchnowski, P., Hunke, M., Büsching, D., Meier, U., and Waibel, A., "Toward Movement-Invariant Automatic Lip-Reading and Speech Recognition", in International Conference on Acoustics, Speech, and Signal Processing, 1995 (ICASSP-95), vol. 1, pp. 109–112, 1995.

[9] Essa, I. A., and Pentland, A., "A Vision System for Observing and Extracting Facial Action Parameters", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 76–83, IEEE, June 1994.

[10] Tamura, S., Iwano, K., and Furui, S., "A Robust Multi-Modal Speech Recognition Method Using Optical-Flow Analysis", in Extended summary of IDS02, (Kloster Irsee, Germany), pp. 2–4, June 2002.

[11] Mase, K., and Pentland, A., "Automatic Lipreading by Optical-Flow Analysis", in Systems and Computers in Japan, vol. 22, pp. 67–76, 1991.

[12] Fleet, D. J., Black, M. J., Yacoob, Y., and Jepson, A. D., "Design and Use of Linear Models for Image Motion Analysis", International Journal of Computer Vision, vol. 36, no. 3, pp. 171–193, 2000.

[13] Martin, A., "Lipreading by Optical Flow Correlation", tech. rep., Compute Science Department University of Central Florida, 1995.

[14] Chițu, A. G., Rothkrantz, L. J. M., Wiggers, P., and Wojdeł, J., "Comparison between different feature extraction techniques for audio-visual", *Journal on Multimodal User Interfaces*, vol. 1, no. 1, pp. 7--20, Springer, March 2007.

[15] Chițu, A. G., and Rothkrantz, L. J. M., "Building a Data Corpus for Audio-Visual Speech Recognition", *Euromedia2007*, ISBN 9789077381328, pp. 88-92, April 2007.

[16] Chițu, A. G., and Rothkrantz, L. J. M., "The Influence of Video Sampling Rate on Lipreading Performance", *12-th International Conference on Speech and Computer (SPECOM'2007)*, ISBN 6-7452-0110-x, pp. 678-684, Moscow State Linguistic University, Moscow, October 2007.

[17] Wojdeł, J.C., Wiggers, P., and Rothkrantz, L.J.M., "An audio-visual corpus for multimodal speech recognition in Dutch language" In *Proceedings of the International Conference on Spoken Language Processing (ICSLP2002)* (Denver CO, USA, September, pp. 1917-1920), 2002.

[18] Varga, A., and Steeneken, H., "Assessment for automatic speech recognition II: NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems." *Speech Communication*, Vol. 12, no. 3, pp. 247-251, July, 1993.
[19] Wiggers, P., Rothkrantz, L. J. M., "Exploring the Influence of Speaker Characteristics on Word Use in a Corpus of Spoken Language using a Data Mining Approach", *Specom 2007*, ISBN 6-7452-0110-x, pp. 633-638, Moscow State Linguistic University, Moscow, October 2007.
[20] Cootes, T.F., Edwards, G.J., Taylor, C.J., Active Appearance Models, In H.Burkhardt and B.Neumann, editors, 5th European Conference on Computer Vision, Vol.2, 484-498, Springer, 1998.
[21] Williams, J. J., Rutledge, J. C., Garsteckiy, D. C., and Katsaggelos, A. K., "Frame rate and viseme analysis for multimedia applications", *Proc. IEEE Works. Multimedia Signal Process,* pp. 13-18, Princeton, 1997.
[22] Williams, J. J., Rutledge, J. C., Aggelos Katsaggelos, K., and Garstecki, D. C., "Frame Rate and Viseme Analysis for Multimedia Applications to Assist Speechreading", *The Journal of VLSI Signal Processing*, Volume 20, Numbers 1-2 / October, 1998, pp. 7-23, November 29, 2004.
[23] Visser, M., Poel, M., and Nijholt, A., "Classifying visemes for automatic lipreading", In *Proceedings of TSD'99,* 1999.

AUTHORS BIOGRAPHY

ALIN GAVRIL CHIȚU was born on November 8, 1978 in Buşteni, Romania. He graduated in 2001 at the Faculty of Mathematics and Computer Science at University of Bucharest, which is one of the top universities in Romania. In 2003 he received the MSc. degree in applied computer science at the same university. Starting September 2003 he joined the Risk and Environmental Master Program at Delft University of Technology, Delft, The Netherlands which he graduated with honors in August 2005. Since then he is pursuing his PhD degree in the Man-Machine Interaction Group, Mediamatics Department at Delft University of Technology under the supervision of Dr. Leon J.M. Rothkrantz. His main interest is in data fusion as the means to build robust and reliable systems, audio-visual speech recognition being one of the case studies. He is also interested in robust computer vision, machine learning and computer graphics.
Email: a.g.chitu@tudelft.nl
Web address: http://mmi.tudelft.nl/~alin

LEON J.M. ROTHKRANTZ received the MSc. degree in mathematics from the University of Utrecht, Utrecht, The Netherlands, in 1971, the Ph.D. degree in mathematics from the University of Amsterdam, Amsterdam, The Netherlands, in 1980, and the MSc. degree in psychology from the University of Leiden, Leiden, The Netherlands, in 1990. He is currently an Associate Professor with the Man-Machine Interaction Group, Mediamatics Department, Delft University of Technology, Delft, The Netherlands, since 1992. His current research focuses on a wide range of the related issues, including lip reading, speech recognition and synthesis, facial expression analysis and synthesis, multimodal information fusion, natural dialogue management, and human affective feedback recognition. The long-range goal of his research is the design and development of natural, context-aware, multimodal man–machine interfaces. Drs. Dr. Rothkrantz is a member of the Program Committee for EUROSIS.
Email: l.j.m.rothkrantz@tudelft.nl
Web address: http://mmi.tudelft.nl/~leon

MEDIA-BASED TRAINING

SIMULATION-BASED TRAINING IN ENGINEERING

Valery Vodovozov
Zoja Raud
Tallinn University of Technology
Ehitajate tee 5
19086 Tallinn, Estonia
zoja@cc.ttu.ee

KEYWORDS
Education and Training Using Simulation, CAI, Synthetic Environments and Distributed Simulation, Teleworking, Electronic Data Interchange.

ABSTRACT

Active learning opens the new possibilities for successful obtaining engineering profession being combined with the traditional approaches. An effective simulation-based approach is proposed in the paper. Education and training using simulation encourages the reinforcement techniques that focus on a conceptual understanding and the new opportunities for students to choose the content and study methods by electronic data exchange. It shows students that teaching is stimulating and caring and gives them time to process the concepts, rather than overworking within the course or curriculum. Additionally, some recommendations are given to improve the assessment.

INTRODUCTION

The Bologna meeting of 1999 set the convergence of all the national university programs of the European partners on a path towards a common frame (Bergan and Rauhvargers 2006). The main recommendations of this meeting concern the equivalent degrees that come into being on more specific subjects and areas of shorter duration. A practical approach to teaching becomes more important in technological degrees, where students graduate and search for jobs in fast evolving companies. To define a common framework for comparing the same studies, the European Credit Transfer System (ECTS) has arisen that measures the work burden of a subject in terms of time a student can do in one hour.

Learning involves approaching of the three human goals. The first goal is a knowledge acquisition, conceptualizing learning as a process of transferring knowledge to an individual leaner. The second is to strengthen the social status, which emphasizes the role of social communities in learning and professional development. The third is a knowledge creation, whose aim is the purposeful generation of information and the development of related social customs.

In this paper, the means applied to achieve these multifaceted goals in engineering education are discussed, as well as the impact of the changes on the student time allotment and the learning paths in technical universities. Here, the current teaching procedures are referred to as *traditional teaching* for comparison with *alternative teaching* approaches discussed in the following sections.

The traditional teaching engineering disciplines, as demonstrated by current textbooks and researches, focuses on the procedural, quantitative and analytical methods to describe common and individual knowledge in the field (Hudson et al. 2008). This technology divides the courses into lectures, labs, practices, homework, and assessments. Most courses follow quite a usual format for the university engineering classes − two or three hours of lecturing and one hour of exercises and problems per week, plus a final exam at the end of a semester.

In the lecture, theoretical concepts are explained and illustrated using the calculated examples. The lecture plan is prepared beforehand by the faculty in accordance with a syllabus and the course duration. With the rapid growth of new technologies, most current textbooks provide more and more space for new devices and circuits. The students' motivation to put effort into their attendance of lectures is appeared low and the lecturers felt that the learning results are poor. The students do not learn well from the lecture format, and their skills cannot develop sufficiently to allow them to solve the related exercises and problems.

There are many drawbacks in these lectures, the primary ones being minimizing or eliminating the focus on intuition. The lecturers feel that the current teaching methods are inadequate and that it is time to reshape the teaching. To improve student confidence with engineering, some professors readopt a traditional lecture by including a conceptual understanding about the behavior of a system in addition to the formal topic explanation. To provide a course environment that encourages a deeper understanding, the lectures are reinforced using visual teaching techniques to reach the dominant sensory learning style. This smart lecture "enrichment" by multimedia poses two immediate problems (Stiles 2000). Firstly, the effort and skills required to produce such content make it unrealistic in terms of both cost and development time to producing a significant body of content across curricula. Secondly, as the learners' expectations of multimedia have been formed by the computer games and films, they will be unimpressed even by relatively expensive multimedia educational content produced by commercial publishers. Therefore the use of multimedia should focus on its value in the learning context, rather than a desire to excite with its "richness".

Laboratory and practical works play an important role in learning greatly increasing students' understanding. The lab classes are the necessary part of a curriculum for engineering courses. In laboratory lessons, the practical systems are implemented based on the theoretical lessons. In addition, such kind of activity develops students towards the following learning goals:

- gaining skills and experience with equipment
- gathering, manipulating, and interpreting data
- cooperation and group work (collaborative learning) − students learn to communicate in order to distribute the workload, discuss problems, and integrate the overall program
- becoming motivated and excited

The students usually carry out these laboratory lessons either individually or in groups of two to four. The latter is the most usual choice. This structure is quite common in engineering education. In a lab class, students work follow a lab manual, set up circuits, test items required by the manuals, and then analyze results.

Nevertheless, the traditional labs and practice suffer from some major disadvantages (Magdalena et al. 2008; Savander-Ranne et al. 2008).

- guided exercises and practical sessions often promote a passive attitude in the student, who may make an effort to fill the spaces in the practical notebook, being uncritical of the work done thus the result is that, frequently, the student carries out the required task, but does not understand the concepts that underlie the work
- another issue is that "parasitism" is possible under this policy as some students in the group can do all the work, while other ones benefit by receiving the same mark, without having done any work at all
- a further drawback is that most of the time, practical sessions do not investigate a practical system but merely work as an example of theoretical concepts, so there is not much additional teaching in these practical lessons
- the current lab procedure has just a repetitive structure built on a mechanical tasks where the components are changed from one form to another, but which bears no relation to the industrial setting
- in addition, this policy promotes neither competitiveness amongst students, nor the ability to assign the tasks.
- the studies of the time use have also shown that students spend far less time studying than are allocated in curricula

The reinforcement techniques (homework and exams) tend to be of the two kinds:

- quantitative, asking students to recognize a standard device and apply the correct formula for the new component values
- procedural, asking students to find the parameters of a new device following the same procedures shown in class

The above given overview of the traditional methods can be summarized as they encourage a *surface approach* to learning, where students try to follow the routine solution procedures and match patterns, rather than a *deep approach* to learning, where students develop a conceptual understanding of how the technical system operates. A surface approach focuses on being able to repeat what has been read without necessarily understanding it. Students' study habits, focusing almost exclusively on exams, lead eventually to an overload, which in turn tends to result in a low level of understanding. This effect is certainly the reason for students' stress and missing motivation. Therefore, the traditional teaching is felt to be ineffective in generating student enthusiasm and passion for learning. In this case, students may not fully achieve the required goals, or may be slow to reach their full creative talent and engineering potential.

The continuous increase in information makes it necessary that the engineering students be able to acquire new knowledge rapidly and process it in situations where change is the norm. It is of utmost importance that the engineering subjects include means by which the students learn to study efficiently and can exercise skills to learn how to learn.

Some recent works brought the novel methods based on the *active learning* strategy. Next sections of this paper analyze the benefits and restrictions of the alternative active learning techniques such as project-based learning and problem-based learning, respectively. Follow the feature description of each method the recommendations are given concerning the possible ways of their use in engineering education. Then, the simulation-based approach proposed by the authors is described and its application is outlined. Finally, the novel assessment approaches are drawn and conclusions are summarized.

PROJECT-BASED LEARNING

Contemporary changes in the business area pose a number of challenges to the present and future electrical engineers, including:

- work in cross-functional teams
- work in multidisciplinary teams
- work in multinational teams
- work in geographically dispersed teams
- working with a global customer base
- developing communication skills
- learning to apply, transfer and improve engineering knowledge

As learning is a cognitive process, knowledge has to be constructed in the mind of the learner's efforts and activities. Learners develop the complex reasoning skills effectively when actively engaged with the material they are studying. They build meaning when read, write and discuss. Learning is aided by a conversation that seeks to describe and clarify the ideas of learners. Clarifying of understanding requires that students have the opportunities to articulate their ideas, to test those ideas through experimentation and discussion, and to consider the connections between them. From this viewpoint, project-based learning seems extremely prospective (Aldrich 2005).

Project-based learning, called also *learning by doing*, is a student-centered strategy that fosters initiative and focuses a student on authentic real-world open projects that can increase motivation for the majority of students and enhance their education (Chu et al. 2008). One aim of this approach is to organize the courses so that the students would be motivated to study regularly, not just for a final exam, and so that they would realize that a deep learning leads to better understanding and, eventually, to a lesser burden. When the total hours allotted to the particular courses are fixed by the curriculum, the only method to introduce the alternative learning proposes is the removing of the prescribed border between lectures and other kinds of study thus allowing students to spend the learning time in other way.

At project-based learning, the style of lectures changes and the behavior (intuitive) analysis replaces the formal explanation of the theory. From now on, practice and labs may precede lectures, if necessary, as well as the new kinds of study appear. Open problems are framed by the teachers in accordance with the topics in the syllabus and a lecturer acts as a contractor. He asks the students to design and implement a fully functional system that covers all the topics reviewed in the theory lessons. During the lessons, the lecturer reviews the designs and products and advises the students about the possible errors or mistakes. Therefore, the teacher is responsible for ensuring that each group's design is correct. The final prototype is usually implemented on prototype boards. Here, *collaborative learning* becomes the predominant method of study for the full-time students.

The following principles underlie project-based learning:

- students are encouraged to work regularly throughout the course and to take an active role in their learning process
- active and reciprocal interaction appears between students and lecturers as well as among the students themselves
- the course content focuses on essentials and emphasizes comprehension of overarching principles
- students are stimulated to obtain relevant concepts and knowledge from preceding courses
- students have to use the new concepts, and conceptual understanding is valued
- teaching of the new content is connected to the students' prior knowledge

Thanks to the project-based approach, the students not only maximize their practical learning experience to achieve the project goals, but also develop other important abilities in:

- *self-directed learning* − as a project may involve multidisciplinary knowledge which is not covered by the standard lecture material, students are driven to study and seek solutions which serve to enhance their understanding of the theoretical material
- *project management* − students organize a task based on the talents of each group member, and each defines their own task and manages their progress against a specified timeline
- *product design* − students have more scope to develop the project so as to display their inventiveness

All of these skills are the critical requirements in engineering education. By providing students with the opportunity to apply the knowledge they have learned, the project-based courses become more appealing to them.

However, in engineering disciplines the traditional studies are required also. The lectures provide the students' understanding of operation principles, give them confidence with the material, and encourage students to take a deep approach to learning. The formal analysis given in the lectures is critical to formulate equations for the system behavior understanding. The feedback from the formal explanation to active experimentation teaches students to trust their formal conclusions and to understand the drawbacks of their answers. Therefore, the courses must contain both the traditional lectures and the self-directed learning material. All of these components are important. Moreover, as part-time and extramural learning usually cannot be involved into the collaborative learning, other methods of the project-oriented work are to be used.

As the goals of engineering education are multifaceted and include the development of skills in such areas as communication and teamwork, writing and presentation, argumentation and debating, training in these skills is to be integrated in the teaching methods and the format of the project-based courses.

Particularly, in Aalborg University over the last 25 years it has been running the project-organized undergraduate education having unique pedagogic model of teaching. In this method, a great part of the semester the teachers' and students' work revolve around the complex real-life problems or issues that the students consider and try to resolve while working together in groups. Half of each semester is devoted to a project and the remaining half to courses (Magdalena et al. 2008).

Similarly, the curriculum structure proposed in (Macias-Guarasa et al. 2006) consists of eight courses: four theoretical courses and four project-based courses (including a compulsory master's thesis). In the project-based courses, the students, working together in groups, develop multidisciplinary systems, which become progressively more complex. An important result is that all students have developed more complex and sophisticated systems, while considering that the results are worth the effort invested.

The same concern the labs. In the complex engineering courses, high mains voltage as well as the mechanical hazards associated with converters and motors make it difficult to introduce the project-based approach into self-learning (Martin and Brown 1998) thus other methods are to be discussed.

PROBLEM-BASED LEARNING

People interpret new experiences and knowledge in terms of particular concepts already present in their memory. Prior knowledge and perceptions influence the understanding and construction of the new concepts. This has a significant impact on how new knowledge is constructed as new

information is integrated with the old and familiar knowledge. For conceptual change to be achieved, the knowledge models require rebuilding. In this process, the learner constructs a new conception or model that better fits the case and then replaces the prior conception. To do this, the learner should be given opportunities to apply the new conception or model in solving scientific problems in order to internalize it and to develop skills using it correctly and scientifically.

Problem-based learning called also *learning by exploring* is the pedagogical approach in which learning is linked to applied research and development projects to encourage students to learn through the structured exploration of a research problem. It means learning expertise that arises from social interaction, knowledge and competence sharing, researching and problem solving of collective objects (Mantri et al. 2008). Reworking the traditional lecture and tutorial models, students usually work in the small self-directed teams to define, carry out and reflect upon a research task, which can often be a real-life problem. The tutor acts here as a facilitator and resource person to whom they can meet for an advice or guidance. The model emphasizes on cooperation and creating a "learning and developing" culture and makes it possible to include and use various scientific perspectives and methods of learning, researching and developing in operation and action.

The problem-based learning process starts by identifying the initial problem or research object, analyzing and describing it, and selecting appropriate work methods. The work represents a continuous problem-solving process, focusing on research, development and generating new competences. The outgoing result is a creation of a novel operating method, a model, a service or a product. The new outcomes may be reached in this way:

- motivation the students to become interested in the subject
- assisting students to master the fundamental concepts
- fostering critical thinking
- forming and testing hypotheses
- development the problem solving strategies and techniques
- mastering research and development methods
- teaching students how to learn and acquire lifelong learning skills
- ensuring that exploring is attuned to the world of work

Having a complete set of the learning objectives is crucial here in determining the direction of the work, acting as incentives both to reach the desired solution and to acquire a range of knowledge and skills while moving towards this solution. In addition, at times, more learning objectives are added while work is in progress. For each problem, the team is encouraged to elect a team leader, who would organize the work distribution. The teams are then encouraged to distribute the work amongst themselves, with team members performing tasks such as searching for information from various resources, compiling the data, doing calculations, performing experiments and finally recording the work done and preparing a presentation. Ideally, these various tasks are rotated among the team members. The teams discuss issues, decide their own theoretical, practical and software goals and explore these learning and mistakes. This encouragement is given by continuous monitoring and by instructing them to record each relevant finding, any mistakes committed and the corrective action taken. Once the team members reach the solution, they are also asked to frame similar kinds of problems and identify application areas. Throughout, they are guided by an instructor who would correct students as they encounter difficulties and help them to draw conclusions to reach the desired goal. They also understand that learning while exploring is as important as finding the correct solution.

During this education, cooperation with the particular researchers and scientific institutions is usually initiated. Together with the lecturer, the researchers plan and introduce new teaching methods, which they anticipated could be used for the next implementation of the course. The researchers make observations during the lectures and give feedback to the instructor on a regular basis, surveyed the students' opinions and prepared questionnaires for the students to fill in and assess the course (Savander-Ranne et al. 2008). The seminars on the self-learning topics are the other alternative teaching method used for learning by exploring.

Problem-based learning is mainly beneficial for the graduate students. These courses are usually targeted towards the last-year master engineering students who are soon to start their final theses. Nevertheless, some of its components may be effectively used in bachelor training.

SIMULATION-BASED APPROACH

An above given review leads to a *strategic approach* to a course design which is output driven and focuses on the effect it has on the learner, rather than an input-led view which focuses on a body of content and its absorption by the learner. This approach can be summarized as:

- learning outcomes should make clear to learners "where they will be" at the end of the course making the context for learning as transparent as possible
- whatever the level of the learners' current tacit and specific knowledge, the educational activities should be realistic and any learning opportunity should be clearly related to the outcomes of the course
- appropriate to the level of the learner, the higher level opportunities can be pre-deconstructed into the lower level opportunities

A *simulation-based approach* proposed further comprises in different proportions all the discussed resources, such as theory (subject-specific, information), internal resources (labs, exercises, discussions, tests), external means (reading lists, web resources, and published lectures), and hints (specific procedural advice and strategic guidance). This kind of learning refers to the activities where students manipulate or build models thus opening the way to integrate the simulation into the traditional lectures and exercises as well as into the real labs any time when such

integration gives more efficiency than the traditional methods do.

The mathematical background of the simulation-based approach is given in (Clement and Rea-Ramirez 2008; Carmel and Markovitch 1998). It was originated from the machine learning techniques that have long been applied in the artificial intelligent field and have gained a lot of success (Jianguo et al. 2007). The purpose of the learning processes is generally to obtain a set of parameters based on a given data set by minimizing a certain objective function which can explain the data set in a maximum likelihood or minimum estimation error sense. However, most of the learned parameters are highly data dependent and rarely reflect the true physical mechanism that is involved in the observation data. In order to obtain the inherent knowledge involved in the observed data, it is necessary to combine physical processes with the learning process rather than only fitting the observations with a black box model (Brian et al. 1998).

The simulation-based approach has been implemented into the educational process at Tallinn University of Technology. The common ideas of this implementation are as follows.

All the educational resources – theoretical, external, internal, and supplemental – have been doubled by the models of different intellectual levels. Particularly, all the lectures, manuals, and exercises are published in the Internet and in the printed materials. The full range of the laboratory works are repeated as the working models operated via the Internet, particularly being involved into the department e-laboratory. The great numbers of experiments that might be executed but could not be supported by the actual labs are prepared as the virtual works along with the consequent their description. The detailed explanation of the possible outcomes of the further activity is published also. The most prospective educational results were obtained from distance learning based on the resources geographically distributed among the partner technical universities and research laboratories.

Using these resources, a learner together with his teachers may develop the individual educational trajectory, following which they reach the specific educational goals combining the traditional, project-based, problem-based and simulation-based activities. During the building of the learning path, the world-wide open resources are involved as the educational components as well, thus expanding the teaching potential of the host institution.

Dependently of the goal of education, multiple educational trajectories may be designed this way. Although the content of the initial curriculum is identical for all the students, each of them has his school degree, interests and the background knowledge. The students' possibilities often motivate them to increase the learning capacity as they progress through the program. The parents, industry, society and teachers have very high expectations from the students. Therefore, continuous refinement, updating and learning new engineering and professional competencies is called for during the years spent at a university for the degree for which a student is registered. The outline of the individual curriculum is agreed with the similar curriculums of other partner participants and further may be edited and improved during the education process.

The significant component of the simulation-based instruction is the original system built to study the Electric Drive discipline.

An electric drive is an electromechanical system that provides controlled inductive conversion of electrical energy into mechanical motion. It includes mechanical transmissions, electric motors, electronic power converters, and informational controllers. The power of supply lines is transformed into motor supply energy by the electronic power converters. Then, the motors convert it into electromagnetic energy, which in turn is transmuted into shaft mechanical energy. The controllers form the input references using information of the set-points, outputs, and disturbances. The engineering goal is to build an effective system using the components proposed by many companies of worldwide industry. Effective assembling of electric drive meets with problems owing to rather complicated algorithms. A list of the professional tasks includes an equipment choice, structural synthesis, experiment planning and its result processing, as well as an implementation and maintenance of the designed systems. Here, the mechanical, electrical, electronic, and power engineering problems are integrated closely.

In accordance with the active learning principles sited above, it is not sufficient just to learn about professional and other issues and to pass exams. The techniques need to be used in real situations by the graduates. For the simulation-based approach, this is extremely important to emphasize the connections between different aspects, to encourage a broad system view and to illustrate the practical, technological and human constraints in solving the real-world problems.

The theoretical component deals with the formal theory about electric drives. Today this theory has reached a degree of complexity that can only be managed by employing computers. Developing the formal models brings forth simulating aspects of the human-machine interaction and implementation as the computer programs. These techniques constitute the basis for the evaluation and further development of the theory. Moreover, the theoretical studies do not always have to be passed before taking the practical ones. To the part of students, it is better to obtain the theoretical knowledge first, but to other it is preferable to pass practices beforehand because this can provide a good basis for the theory thus helping students to understand some of the theoretical concepts. Consequently, multiple educational trajectories may be built that depend on the students' abilities and course contents as well as the different ratio of the lecturing, practical and assessment lessons are possible.

The specially prepared textbook (Vodovozov and Vinnikov 2008) and its web release involve the sections supported the simulation-based learning. Two of them are destined to accompany the lectures for beginners and for advanced

learners summarizing the common terms and conditions as well as the topical mathematical basing of the course. A number of specific sections explain the broad self-learning part of the course. Other sections are devoted to the calculation examples, experimental and assessment problems of the course. The multiple links connect the textbook with the analogous and supplemental works in the field. The properly structured index and the reference list serve as the powerful navigation tool.

The laboratory equipped with the help of the leading firms in the field supports the theoretical course. It includes a number of test benches that provide the e-learning on-site and off-site activities (Vinnikov et al. 2008).

The specific toolbox *eDrive* developed to support the course in accordance with the project-based and problem-based requirements as well as for the assessment goals opens multiple possibilities for learners and instructors (Vodovozov and Raud 2008). The main features of an educational soft tool that discern it from the models meant for the professional designers are as follows:

- descriptiveness and compatibility of results
- clearness of physical essence
- suitability of report generation and format conversion
- independence of particular company interests
- matching the standards and design rules
- availability of learning-oriented manuals, textbooks, and guides

The described software offers the solution of the next typical educational and project management problems:

- informational support in choosing the electrical, mechanical, and electronic equipment
- mathematical and computer simulation and full computation with the use of databases
- testing and result verifying in accordance with different criteria
- the drive tuning and the load optimization

The toolbox includes the next components:

- the powerful database
- a set of adjustable controller schemes
- the models of motors, converters, and gears
- the graphic package for representation of the steady-state and dynamic simulated processes with an automatic and manual scaling, report generator, system analyzer, and preview facilities
- the signal generator to produce the test reference and load signals as well as the nonlinear curves, noses, and filters

Thanks to the powerful mathematical core, it supports all the laboratory works of the department. Moreover, the environment provides the possibility to simulate the same works using the equipment of a number of companies that cannot be presented in the institution. Comparing and discussion of the obtained results serve as a very beneficial instrument in the development of the students' skill and experience.

To stimulate the teamwork, the specific methodology has been developed (Vodovozov, 1995). The goal of the collaborative design is to build an effective electric drive of a real-world machine – robot, roll-table, crane, electrical car, etc., using the equipment databases of the leading companies in the field and their design approaches. Each student asks to design his own drive component and has to develop it follow the common technical rules and industry standards that the machine is to be fulfilled.

To support research and development learning, a series of exploring problems is proposed. For this, some international seminars for doctoral students and the master school were arranged in conjunction with the *eDrive* methodology. In such environment, some new distance learning activities were tested and discussed.

ASSESSMENT IN ACTIVE LEARNING

To define a common framework for comparing the same studies, the European Credit Transfer System (ECTS) exists that measures the work burden of a subject in terms of time a student can do in one hour (Bergan and A. Rauhvargers 2006). In traditional learning, the grading scheme is largely prescribed by the host university. This evaluation division for the examinations and that for the practical credits is usually given in the curriculum. The students are required to take the theory exams, as this exam serves to qualify them for the next semester.

Modes of assessment can often also be criticized for being too much oriented towards the exams, with very few other forms of evaluation and feedback being used. In laboratory estimation process, the questions posed to students regarding the important aspects of the work typically give a subjective and narrow mark. Therefore, the evaluation strategy has to be redefined and reformulated for the goals of active learning and additional assessment methods are required to stimulate a learner. The students' assessment is needed to receive currently the actual feedback (Savander-Ranne et al. 2008).

Particularly, the assessment in simulation-based learning invokes to evaluate:

- the problem statement and understanding
- the learning objectives
- the methodology used
- the problem solution under the theoretical and practical headings
- calculations, simulation execution, and software selected
- practical results, presentations, and the printouts prepared

To activate the learners' activity, in the first lecture the course goals are discussed, the rules are set, and the assessment criteria are introduced, which are mutually agreed upon by the students and lecturer. All tasks are described and their influence on the final grade is explained and justified. This information is available on the course's Internet pages, where all other relevant materials are also posted as the course progressed.

To make lectures more appealing to students during the semester, the pre-lecture assignments and concept tests are typically used regularly. For this, at the beginning of the most of the lectures, the students take a concept test dealt with issues from the preceding lectures. The tutor may give 8 to 12 assignments, each containing 15 to 25 questions to students. Students are asked to find any one to four right answers to all questions from each question form. Answering the concept test takes usually 5 to 10 minutes. Immediately afterwards, the lecturer discusses the main concept test problems. This discussion covers both the correct and the incorrect responses to the alternatives of the multiple-choice questions, and gives thorough rationale and justification. This feedback discussion on the concept test needs sometimes 5 to 10 minutes. Sometimes, at the end of a lecture a summing-up discussion stimulates active learning again. Also, it gives the constructive feedback to the students involved.

So called "short answer" conceptual questions proposed in (Hudson et al. 2008) serve as an effective addition to traditional quantitative analysis questions on exams and homework.

As concerned the collaborative methods, the learning assessment is devoted to examine the nature and appropriateness of the students' collaborative and group working potential. In simulation-based learning, in accordance with recommendations of (Magdalena et al. 2008) the practical mark counts towards 25% of the student's total grade, with the other 75% being for the final exam that every student passes at the end of the semester. A student's mark for this practical part is decided basing on the marks obtained for his project-based or problem-based activity, such as the cost of the prototype developed, specifications achieved, the final report and the public presentation. Therefore, the variety of assessment methods is very broad and their efficiency may be enough high.

CONCLUSION

Students who take an active approach to learning reach a higher, more integrative level of understanding and demonstrate longer retention of the material. Technical graduates using a strategic approach to learning have a strong inclination to apply the type of learning that forces passing the courses. Since a student's orientation to learning can be modified by the course environment, instructors have the ability to influence which approach a student will adopt. The proposed simulation-based instruction encourages the reinforcement techniques that focus on a conceptual understanding and the new opportunities for students to choose the content and the study methods. It shows students that teaching is stimulating and caring and gives them time to process the concepts, rather than overworking within the course or curriculum. The experience of Tallinn University of Technology demonstrates that reinforcing the simulation-based technology will result in more students' adopting of active learning.

ACKNOWLEDGEMENT

This research was supported by European Social Fund for Doctoral Studies and International Programme DoRa.

REFERENCES

Aldrich, C. 2005. *Learning by Doing: A Comprehensive Guide to Simulations, Computer Games, and Pedagogy in E-Learning and Other Educational Experiences*, J. Wiley & Sons, 356 p.

Bergan, S. and A. Rauhvargers (Eds.). 2006. *Recognition in the Bologna Process: Policy Development and the Road to Good Practice,* Council of Europe Higher Education, ser. 4, 202 p.

Brian C.; B. C. Williams; and W. Millar. 1998. "Decompositional Model-Based Learning and its Analogy to Diagnosis", *50th National Conference on Artificial Intelligence − Innovative Applications of Artificial Intelligence*, Madison: WS, pp. 197-204

Carmel, D. and S. Markovitch. 1998. "Model-Based Learning of Interaction Strategies in Multi-Agent Systems", *Journal of Experimental and Theoretical Artificial Intelligence*, 10, pp. 309-332

Chu, R. H.; D. D. Lu; and S. Sathiakumar. 2008. "Project-Based Lab Teaching for Power Electronics and Drives", *IEEE Transactions on Education*, vol. 51, no. 1, pp. 108-113

Clement, J. J. and M. A. Rea-Ramirez (Eds.). 2008. *Model-Based Learning and Instruction in Science*, London: Springer, 286 p.

Hudson, T. A.; M. Goldman; and S. M. Sexton. 2008. "Using Behavioural Analysis to Improve Student Confidence with Analog Circuits", *IEEE Transactions on Education*, vol. 51, no. 3, pp. 370-377

Jianguo, W, J.; X. Lu; and J. Dang. 2007. "A Model-Based Learning Process for Modeling Co-Articulation of Human Speech", *IEICE Transactions on Information and Systems*, E90-D(10), pp. 1582-1591

Macias-Guarasa, J., J.M. Montero, R. San-Segundo, A. Araujo and O. Nieto-Taladriz. 2006. "A Project-Based Learning Approach to Design Electronic Systems Curricula", *IEEE Transactions on Education*, vol. 49, no. 3, pp. 389-397

Magdalena, R.; A. J. Serrano; J. D. Martin-Guerrero; A. Rosado; and M. Martinez. 2008. "A Teaching Laboratory in Analog Electronics: Changes to Address the Bologna Requirements", *IEEE Transactions on Education,* vol. 51, no. 4, pp. 456-460

Mantri, A.; S. Dutt; J. P. Gupta; and M. Chitkara. 2008. "Design and Evaluation of a PBL-Based Course in Analog Electronics", *IEEE Transactions on Education,* vol. 51, no. 4, pp. 432-438

Martin, T. W. and W. D. Brown, 1998. "Experiences With a Laboratory-Intensive Curriculum in Electrical Engineering", *28th Annual Frontiers in Education Conference,* (*FIE '98*), vol. 3, pp. 1145-1148

Savander-Ranne, C.; O.-P. Lunden; and S. Kolari. 2008. "An Alternative Teaching Method for Electrical Engineering Courses", *IEEE Transactions on Education,* vol. 51, no. 4, pp. 423-431

Stiles, M. J. 2000. "Effective Learning and the Virtual Learning Environment", *Towards Virtual Universities* (EUNIS 2000), Instytut Informatyki Politechniki Poznanskiej, Poznan.

Vinnikov, D.; T. Jalakas; I. Roasto; Z. Raud; and M. Egorov. 2008. "Versatile Laboratory Tools for Advanced Course of Power Electronics", *11th Biennial Baltic Electronics Conference* (*BEC'2008*), Tallinn, Estonia, pp. 277-280.

Vodovozov, V. and D. Vinnikov. 2008. *Electronic Systems of Motor Drive*, Tallinn: TUT Publishing, 248 p.

Vodovozov, V. and Z. Raud. 2008. "Model-Based Learning in Electric Drive Design", *The 2008 European Simulation and Modelling Conference* (*ESM'2008*), Le Havre, France, pp. 355-359

Vodovozov, V. 1995. "The Educational Resources of Mechatronics", *Mechatronics*, vol. 5, no. 1, pp. 15-24

Personalisation of E-learning courses

Wouter Hustinx [1] , Jeanne Schreurs[2] and George Sammour[2]
[1]PHL University College
PHL, Elfde-Liniestraat 24 - 3500
Hasselt, Belgium
E-mail: Wouter.Hustinx@phl.be

[2]Hasselt University
Hasselt University, Diepenbeek Campus
Agoralaan gebouw D, B3590 Diepenbeek, Belgium
Jeanne.Schreurs@uhasselt.be,
george.sammour@uhasselt.be

KEYWORDS
E-Learning; learning styles; personalisation; learning object

ABSTRACT

The degree to which learners are eager to learn from, and with ICT may be influenced by a number of factors, such as their learning style, prior domain-specific knowledge, computer skills, personal beliefs on ICT, age, et cetera. Because of this multidimensionality, you never know for sure whether or not students will appreciate an e-learning course. But, a designer can make efforts to make the fit as good as possible. Focus is on the design of varied learning paths by the instructor to fit the learning styles of the learners to their characteristics, especially their personal learning styles. Based on authors Kolb, Smith and Vermunt, multiple typologies of learning styles have been introduced. As an illustration, two types of e-learning courses are analyzed on their ability to fit with the personal needs of all learner. In a second model the learning content is becoming more decentralised, can be get from different sources and fits more the characteristics of the learner on point of foreknowledge and interest. However, to realise an optimal fit between e-learning design and individual characteristics of the learners, designers face the challenge to create 'intelligent' e-learning courses.
We argue that these intelligent, semantic types of e-learning with a dynamic generation of learning content, deliver content in a most personalised way.

INTRODUCATION

E-learning is still challenging

The introduction of new technologies in education enables teachers to continually rethink and innovate their lessons. The discovery of new tools offer opportunities to improve the quality of the teaching learning processes. However, a problem many teachers encounter is the number and diversity of the tools available on the internet. A good illustration can be found on the website *www.go2web20.net*, an inventory of web2.0 tools. Teachers who visit this website start wondering how they can implement these tools in their own practice, and why they should. Too often they lack the adequate framework to validate the quality of these ICT-tools. The question mostly being asked is 'what can I do with these tools that I cannot do without them?'
Organizations install a virtual learning environment for shaping their technology enhanced teaching and learning activities. A virtual learning environment is a software system designed to support teaching and learning (Weller, 2007). The platform supports classroom teaching by organizing the electronic course materials. When it comes to distance and blended learning, the virtual learning environment becomes more important, it is at the heart of all learning activities. Blended learning is an example of a learning design, in which we combine e-learning with traditional classroom learning activities. It is straightforward to structure the distance learning activities in learning units. A learning unit is a logical sequence of learning activities seen as sequential steps the learner has to take.
Designing, developing, and deploying e-learning resources are only part of the e-learning battle. Actually getting employees and prospective students and instructors to use e-learning is a challenge on its own. From an organizational managerial point of view, only structural measures are effective for the organization-wide implementation of e-learning. An example of a strategic action is the systematic reduction of

face-to-face classes, replaced by distance learning activities.

Problem of the adoption of e-learning solutions in life long learning

Designing, developing, and deploying e-learning resources are only part of the e-learning battle. Actually getting employees and prospective students and instructors to use e-learning is a challenge on its own. Generally, when discussing e-learning adoption, it is interesting to see that the technology itself is secondary to the situations and needs of the client or learner. Rarely is the discussion focused exclusively on technology. Most often, the question being asked is "what can I do with this that I cannot do without it". In e-learning, the answer usually centers on access, cost, convenience, and learning effectiveness. Following (http://business.cisco.com) are some discussions from e-learning practice:

1. "It takes a lot of explaining, hand-holding, and demonstrating to convince a low-tech person that the flexibility and convenience of distance education are worth the trouble of learning a new trick (our average student age is 37!)"
2. It's not enough to simply convert classroom materials into HTML files and post them to a Web site. E-learning must also be flexible and offer content choices to account for different learning styles of employees.
3. "E-learning must also be compelling, interactive, engaging, and offer feedback options," says Hallett. "In many cases, it means reading less online and providing more audio or video options."
4. Any successful e-learning effort must be compelling from the opening sequence, according to Lisa Sass. "We had to make it visually stimulating and engage the senses as much as possible," says Sass. "We wanted to grab them the way a movie trailer grabs them—right at the very beginning—or they would not continue."

INDIVIDUAL LEARNERS AND LEARNING STYLES

Characteristics of the individual learner

Attention should be paid to the distinctiveness and the individual needs of the *learners*. Each learner has his own, unique characteristics. Instructors must take into consideration these differences among learners. An important question, however, is to what extent a teacher must aim at the individualization of learning paths. In a classroom setting, it is impossible to create a different learning path for each individual learner.
The degree to which learners are eager to learn from, and with ICT may be influenced by a number of factors, such as learning style, prior domain-specific knowledge, computer skills, personal beliefs on ICT, age, et cetera. Because of this multidimensionality, you never know for sure whether or not students will appreciate an e-learning course. But, a designer can make efforts to make the fit as good as possible.

Learning Styles

A *learning style* is a student's consistent way of responding to and using stimuli in the context of learning. Kolb (Kolb, 1984 in Smith, 2001) identified four learning styles: activists, reflectors, theorists and pragmatists. Individuals capable of using all of these learning styles are considered to be well-balanced effective learners (Table 1). In practice, most individuals have one or two preferred learning styles. The other styles are rarely used and they master them to a lower degree.

Table 1 – Inventory of learning styles (Kolb, 1984 in Smith, 2001)

Activists	- Enthusiast for new content - Lose patience quickly - Learn best in competitive and challenging situations
Theorists	- Consider multiple alternatives - Found on a logical model / theory
Reflectors	- Considering before acting - Learn best when given time to prepare
Pragmatists	- Experiment with new plans - Don't like too much reflection or discussion - Learn best when there are clear goals and obvious advantages

We can say that each student learns best using a learning strategy or method that best matches his or her need. Or we can say that what matters the most is the learning process, not the style. What

is the truth? Till to now, not their learning styles but achieving a solid learning environment that meets the student's need, seems the means for effective learning.

Learning styles analysed in the previous section were discussed from the learner's point of view, other typologies of learning styles exist. In particular *the VAK-model* is applicable to the presentation of e-learning content of a self-paced e-learning course (VARK, 2005). The strengths of the VAK-model are its simplicity and its recognisability. Standing in the world means gathering information, stimuli through all our senses. The VAK-model focuses on three main sensory receivers: *Vision, Auditory and Kinaesthetic (movement)*. Learners use all three components to receive information, but normally one of these is dominant. This dominant style determines a person's best way of learning new information. The dominant style may not be the same for each type of information. This means a learner may learn in a different way, depending on the type of learning task.

- ***Visual learners*** can be divided into two subgroups: *linguistic and spatial*. *Visual-linguistic learners* like to learn through written language, such as reading and writing tasks.
- ***Auditory learners*** easily absorb information they hear. They may experience difficulties with reading and writing tasks.
- ***The kinaesthetic learner*** prefers to learn through workshops, practical classes and hands-on activities.

Vermunt (1996) created another classication of learning styles, integrating four components of learning: processing strategies, regulation strategies, mental models of learning, and learning orientations. Vermunt identified four different learning styles: *meaning-directed, reproduction-directed, application-directed* and *undirected.*

In creating a learning module, teachers have to take into account the characteristic of the learners. Many forms of learner activities do exist. Creating a mix of several kinds of learning activities is the main point in the design of learning processes. The variety of learning activities meets more than one learning style. The 'best' learning style does not exist, it is always a matter of situating the learning style into the actual learning context. Thus, it is possible that the best learner in classroom teaching, achieves low grades in distance learning activities, because the latter addresses other skills and requires other attitudes from the learner.

Organisation of a mix of learning activities

It is important to note that besides learning styles, learners also differ in expert level and self-regulation strategies. Learners can be at a novice, advanced or expert level in a content domain. When teaching novices, one must pay a lot of attention to domain-specific knowledge like terminology and patterns. In contrast, when teaching experts, problem solving and knowledge-building are the main objectives. Self-regulative strategies make a learner manage his/her thinking process. This management involves a profound understanding of the problem, (re)planning, monitoring one's progress, making and discussing decisions, and evaluating learning progress (Kayashima, et al., 2003).

Learning can be optimised when the instructor cares about the learning styles of the learners. The selection and presentation of the content must correspond to the learners' characteristics such as prior knowledge, self-efficacy, and learning styles. As we have argued, in each classroom, there is a mixture of learning styles. As designer, it is strongly advisable to present information appealing to different learning styles. This provides all learners, no matter what their preferred style is, with the opportunity to become involved. It also allows a learner to be exposed to the other learning styles. Table 2 shows examples of activities performed by different types of learners (VAK-model). The table could be expanded with Kolb's and Vermunt's learning styles.

DESIGN OF VARIED LEARNING PATHS

Design of the learning process

In traditional classroom settings, the instructor provides learners with content that has to be processed by them. The instructor tries to create the best fit between the learning activities and the characteristics of the learners. During the learning process, learners provide feedback to the instructor in various ways: asking for help, frowning one's brows, becoming noisy, et cetera. The teacher responds immediately by providing guidance or presenting the content in a different way.

Table 2: Learning activities performed by different types of learners

	e-learning activity	**Other e-learning activities-blended learning activities**
Auditory learner	In the beginning of the course the new material has been delivered with a brief explanation of what is coming. At the end a summary of what has been covered concludes the course.	- Questioning of the learners to draw as much information from them as possible and then fill in the gaps with some own expertise. - Include auditory activities, such as brainstorming. - Leaving plenty of time to debrief activities, so the learners can make connections of what they learned and how it applies to their situation.
Visual learner	- Use graphs, charts, illustrations, or other visual aids. - Include outlines, agendas, handouts, etc. for reading and taking notes	- Including plenty of content in handouts to reread after the learning session. - Leave white space in handouts for note taking. - Supplement textual information with illustrations whenever possible. - Have them draw pictures in the margins. - Show diagrams and then explain them.
Kinesthetic learner	- Use activities that get the learners up and moving. - Play music, when appropriate, during activities. - Use colored markers to emphasize key points on flipcharts or white boards	- Give frequent stretch breaks (brain breaks). - Provide toys such as Koosh balls and Play-Dough to give them something to do with their hands. -Provide high lighters, colored pens and/or pencils - Have them transfer information from the text to another medium such as a keyboard or a tablet.

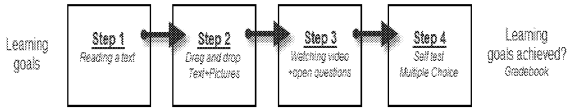
Typical in e-learning is the choice of learners when they take the course and the pace in which they proceed. Communication mostly happens asynchronous, that is, the learner's feedback is not obvious at once for the designer of the learning unit. In this case, the flexible adjustment of the e-learning process is much harder.

In e-learning modules, the structuring of the learning process becomes even more important in comparison to traditional classroom teaching. The teacher seeks to organize learning activities and to deliver learning content in a logical sequence. The realisation of an efficient and effective learning process is the main goal of the designer of the e-learning module. In other words, learners should be able to proceed easily through the e-learning module, and the postulated learning goals have to be achieved by all learners.
The virtual learning environment is a great instrument for teachers to organize, manage and deliver electronic course materials. From a didactic point of view, the combination of various multimedia tools and e-learning activities makes the learning process more attractive for students, and it addresses multiple learning styles. Teachers can provide students with a variety of learning objects they cannot all show in the classroom due to lack of time or poor infrastructure.
We illustrate the method of designing e-learning with two examples.

E-learning course design: E-learning module A

Figure 1 illustrates that e-learning *module A* consists of four steps in a highly pre-imposed structure. In each step, the learner must undertake well-defined tasks individually. The main goal in this e-learning module is the transmission of knowledge.

Figure 1: Sequence of e-learning module A



Learning tasks focus on factual recall and understanding of assigned reading. In this type of learning modules, all learning goals can be

formulated at a very precise level, and relevant content can be collected *(learning content = centralized)*. At the end of the learning process, what is learned can be exactly described.
Teacher guidance is very strong in this type of e-learning, both at the content level and at the formal level. The teacher owns and determines the learning space (*task ownership*), and the teacher decides who does what *(task control)*. Because of this strictly pre-imposed structure, learners' progress can be checked at any time in the learning process. The grading however is focused on the product and thus can be done very easily.
The individual learner is forced to undergo the structure of the learning unit. This type of e-learning modules is suitable for novices in a content domain, and for learners with a low degree of self-regulation strategies.
When we take a look at the learning styles of the VAK-model described in the section above, this learning unit serves both the visual *(text, video,...)*, auditory *(video)* and kinaesthetic *(drag and drop)* learners.
E-learning module A can be criticized because of its very narrow, closed structure. There hardly is any freedom for the learners. According to the design levels revealed by Figure 2, we notice that particularly novice learners will benefit from this highly pre-imposed structure since they know very clearly what they are expected to do. We conclude that this type of e-learning is traditional and only suitable in some situations. In e-learning module B a different approach to e-learning is illustrated.

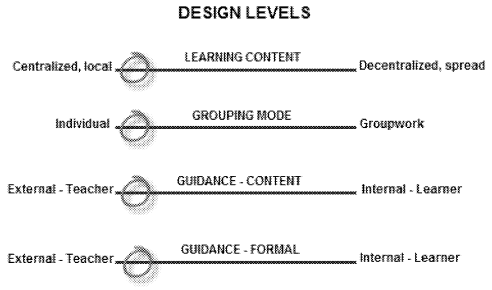


Figure 2: design levels of e-learning module A

E-learning course design: E-learning module B

E-learning module B is based upon the principles of *computer supported collaborative learning (CSCL)*. It fits better with modern theories and insights:

- *Social constructivism:* the active learner is central in the e-learning course, and constructs knowledge in social interaction with others starting from authentic and meaningful tasks.
- *Web 2.0:* the Internet has become a platform which allows anyone to create, upload and share information. Web 2.0 is an umbrella-term for developing social software applications such as wiki's, weblogs, social bookmarking, etc. that facilitate creativity, communication, collaboration, and sharing among users.

- *Homo zappiens:* this concept refers to learners with an active attitude towards diffused information, scanning of computer screens, and using non-linear ways to generate knowledge (Veen, 2000).

In this second example of an e-learning module (see Table 3) more complex learning tasks have to take place. Most tasks are based upon interaction and collaboration with peers in a small group under shared responsibility *(grouping mode: groupwork)*. Higher quality of understanding is likely to emerge because of learners' discussions on the content, preparation of shared products, and selection of relevant materials (cognitive skills), providing each other with feedback, accounting for choices made, et cetera.
The learning objectives of this course are rather complex, open, and hard to define. The external guidance on the content level is minimal: the teacher and each subgroup of learners select the research topic by mutual agreement. The learners are responsible for the interpretation of the content. The teacher, however, has to structure the learning environment in a very detailed way in terms of task instruction, the order of learning tasks, wiki structure, imposed deadlines, available time, etc. *(formal guidance = external, teacher)*. This formal guidance helps learners to complete the learning module successfully.
During the learning process, the teacher performs the role of a coach who enhances learning by creating cognitive conflicts, and who assists the learners into the zone of proximal development. The e-learning designer has searched for a way to focus the assessment on both process and product.
In the year 2008, e-learning module B was developed and implemented in the final (3rd) year of a teacher training program in the PHL University college in Belgium (www.phl.be). At

PHL each student has a personal laptop at his disposal, and can connect wireless to the internet all over the campus. The course, in which this design was initiated, is titled *'Current Educational Issues'*. Some examples of topics that students had to investigate in small groups were, for example, *the problem of school skippers, the debate of knowledge building versus skill development, the amount of culture in curricula.* The technical computer expertise of students was high, since they work with their computers for both learning and personal activities on a daily basis. Along the learning process, it has become clear that teacher guidance at the formal level is essential for successful completion of the module.

Table 3 sequence of e-learning module B

Weeks	Learning activities	Grouping mode	Teachers' guidance	Assessment
1	Exploration of the theme by online brainstorm with videoconferencing software	Small group	Theme selected by teacher *(each group researches a different theme).* Submission deadline at the end of the week.	Electronic group mindmap
2-3	Internet search + readings Active participation in the group discussion area	Individual	Making controversial stands Follow-up of the discussion Defining the role of moderator	Personal bookmark list Minimum of threads per person
4-5-6	Creation of a wiki about the theme	Small group	Defining the global structure of the wiki Defining the role of SPOC Weekly briefing	Critical analysis of wiki's content Group process
7	Presentation of the theme to the other groups (Face-to-face)	Group presents to entire class	Defining presentation criteria Defining peer assessment criteria	Evaluation of the presentation Peer assessment

Furthermore, students tended to look for confirmation from the teacher. For the responsible teacher, it was striking that even advanced students with a considerable amount of self-regulative and collaborative skills needed such an amount of guidance. On the other hand, this guidance enforced and motivated them to accomplish the learning tasks at a high-quality level.

The design fits the personal characteristics of the learner

E-learning module B reflects all the learning styles defined by Kolb. The open nature of the learning module enables each learner to make choices in learning activities and strategies. Theorists and reflectors may benefit more from this design, because of the reading activities and discussions that have to be done.
To create congruence between e-learning and the principles of modern education, e-learning designers must have the competence to adapt the e-learning modules to the personal needs of their students, and to the characteristics of constructivist learning. Additionally, designers feel the pressure to integrate the newest ICT-tools in their e-learning modules. However, it is important to keep in mind that technology itself is secondary to the situation and needs of the learner, as well as to the formulated learning objectives.
The examples analyzed above have demonstrated that a designer can take the individual differences among learners into consideration by creating a varied learning path. External guidance at the formal, structural level is high in both (and most) of the e-learning designs. As Figure 3 suggests that in this model the learning content is more decentralized which means more personalization, it also stimulates group work, finally the learner has the ability to guide himself during the learning process.

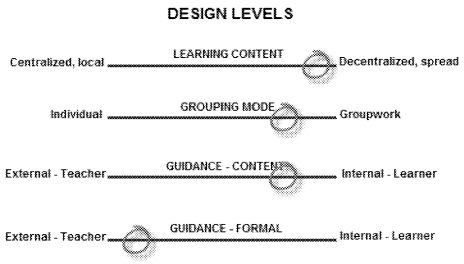


Figure 3: design levels of e-learning module B

Because of the variety of learning styles, a learner will find one step more convenient or easier to take than other steps. The achievement of learning goals is dependent on both the learning path, and the individual characteristics of the learner. Consequently, this results in a

diverse group of learners who achieved the same learning goals but at a different mastery level. 'Smart' learners can advance even further by dividing the learning tasks between group members; hence they only have to accomplish these tasks that best fit their learning styles.
E-learning designers are confronted with the challenge to create e-learning modules that fully correspond to the learning characteristics of the individual learners, yet at the same time accomplish all postulated learning objectives. Instead of trying to focus on each learner's personal style, it is perhaps more important to build an adaptable learning environment. Materials can be presented through a variety of methods, using different ICT-tools, and addressing a combination of learning styles, preferences, prior knowledge, and self-regulation skills.

PERSONALISATION OF THE LEARNING CONTENT

In model B the suggestion is made to personalise the learning content. Learners are all different with respect to learning style, to prior domain-specific knowledge, to computer skills, to personal motivation to use ICT, and can use different kind of appliances to access the learning content. The learning content and the presentation of it has to fit those learner characteristics.

Learning content organised as LO's

The instructor can build LO's based on the original course materials. Those have to be decomposed in small chunks of content. Lo's are a composition of information blocks being composed of the raw materials being the atomic learning objects (ALO) including pictures, summary, key terms, questions. The information blocks can be assembled into LO by adding context and metadata to them. An e-learning module/ course can be created by packaging LO's. Since LO's , defined according to IEEE standards, as an entity that can be used, re-used or referenced during technology supported learning (4)
e-courses can be created by assembling LO's from different sources. From technological point of view, to communicate these materials over the web, it must be based on XML.

Application of semantic web technology

E-Learning systems and e-learning research areas can benefit from semantic web technologies. The Semantic Web technology has enabled by a set of suitable agents, which seems to be powerful enough to satisfy the e-learning requirements fast, just-in-time and relevant learning. This allows for dynamic generation of course content according to the learner preferences, this includes:

- *Distribution:* using LOs *that* will be as decentralized as possible. This enables an effective co-operative content management.
- *Personalization:* A user (using personalized agent) searches for learning material customized for her/his needs. The ontology is the link between user needs and characteristics of the learning material.
- *Dynamic:* Enables the use of knowledge provided in various forms, by a semantical annotation of content. This distributed nature of the LOs enables continuous improvement of learning materials.

Dynamic generation of personalised learning content and e-learning courses

When the learner will log-in to the e-learning system, a set of user characteristics will be loaded for that user logging in, which includes, his age, preferred devices to be used, fonts size, level of courses he has finished in the context of that learning course he is enrolled in. Then the system will dynamically generate the course/learning material according to the user characteristics and preferences, including the layout relevant to his preferred device he is using to view the course content. This also allows more flexibility and efficiency using the system, such that the system will decide the size and format of the content to be delivered to the user according to the device he is currently using, so if he is using a PDA or a mobile device less pictures and videos will be included in the material, for faster communication and loading of content.
The deliveries of these personalised e-learning materials are based on semantic web technology.
Not only metadata is required, but ontologies have to be built to define the relations between the materials parts.
Classification of the content objects is based on RDF and this will optimise the search for relevant content and selection of it. To this end rules have to be followed that are built as a logical structured rule base.

CONCLUSION

We set forward that learners need a high level of personalisation in the delivery process or presentation as well as in the content itself.
Personalisation means the fit with the personal characteristics of the learner, being learning style, prior domain-specific knowledge, computer skills, personal beliefs on ICT, age, et cetera.
Focus was on the design of varied learning paths by the instructor to fit the learning styles of the learners to their characteristics, especially their personal learning styles.
In a second model also the learning content is becoming decentralised and can be get from different sources and fitting the characteristics of the learner on point of foreknowledge and interest. The content has to be organised as learning objects (LO) and following the international standards. To realise a dynamic generation of personalised content into the learning process of the learner we have to base the system on semantic web technology. We outlined the layers of this model and their functions. Our status of our solution is still limited to metadata and the XML format of the LO's.
The next step will be the definition of the ontology and the development of the logical rules to optimise the search and selection of the relevant content, to be generated on real time when the learners enters into the course and fitting his characteristics and his preferences. Part of the content will be organised in the internal learning centre server and part will be get from other and external learning servers.

REFERENCES

Blas, T.M. & Fernández, A.S. (2009). The role of new technologies in the learning process: Moodle as a teaching tool in Physics. *Computers & Education 52 (1)*, 35–44.

De Freitas, S. & Neumann, T. (2009). The use of 'exploratory learning' for supporting immersive learning in virtual environments. *Computers & Education 52 (2)*, 343–352.

J.Schreurs, R.Moreau (2006). "Learning objects aligning different learning styles", Proceedings of the international conference on e-learning. ICEL2006.University of Quebec Montreal. P.415-422. ISBN 1-905305-31-5.

Kayashima, M. & Inaba, A. (2003). *How do we facilitate development of learners' self-regulation skills?* http://www.ei.sanken.osaka-u.ac.jp/~ina/Doc/SR-Skill/aied03kaya.pdf

Lockhorst, D. (2004). *Design principles for a CSCL Environment in Teacher Training.* IVLOS Series.

Schiaffino, S., Garcia, P. & Amandi, A. (2008). eTeacher: Providing personalized assistance to e-learning students. *Computers & Education 51 (4)*, 1744–1754.

N.Friesen(2001) "What are educational objects?"; Interactive Learning Environments, 9.3; p.219-230.

Shaw, G. & Marlow, N. (1999). The role of student learning styles, gender, attitudes and perceptions on information and communication technology assisted learning. *Computers & Education 33 (4)*, 223-234.

Smith, M. K. (2001). David A. Kolb on experiential learning. *The encyclopedia of informal education,* http://www.infed.org/b-explrn.htm.

VARK (2005). A guide to Learning Styles. http://www.varklearn.com/english/page.asp?p=questionnaire.

Veen, W. (2000). *Flexibel onderwijs voor nieuwe generaties studenten.* Oratie. *[Flexible education for new generations of students.* Inaugural speech]. December 15, 2000, Delft, The Netherlands.

Vermunt, J.D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education 31*, 25-50.

Weller, M. (2007). *Virtual learning environments: Using, choosing and developing your VLE.* London: Routledge.

http://business.cisco.com

http://www.go2web20.net

BIOGRAPHY

WOUTER HUSLINX is a teacher educator at the PHL University College, Department of Education and currently does research on the integration of multimedia in education.

JEANNE SCHREURS is Professor at the Hasselt University, Faculty of Applied Economics Campus Diepenbeek, Agoralaan gebouw D, 3590 Diepenbeek, Belgium.

GEORGE SAMMOUR is a PhD student at the Transportation Research Institute — Hasselt University, Diepenbeek, Agoralaan gebouw D, 3590 Diepenbeek, Belgium.

E-BUSINESS AND GOVERNMENT

BUSINESS VALUE SCENARIOS OF INHERENT NETWORK MANAGEMENT APPROACHES FOR THE FUTURE INTERNET

Gerhard Haßlinger
T-Systems Enterprise Services GmbH
Deutsche-Telekom-Allee 7
D-64295 Darmstadt, Germany
E-mail: gerhard.hasslinger@telekom.de

Frank-Uwe Andersen
Nokia Siemens Networks GmbH & Co. KG
Siemensdamm 62
D-13627 Berlin, Germany
E-mail: frank-uwe.andersen@nsn.com

ABSTRACT

Today, the management of broadband access networks is usually centralized with simple support functions in the network elements for monitoring and triggering alarms in failure cases. The future Internet will connect heterogeneous parts for different fixed, wireless and mobile transmission technologies, as well as overlays on network and service layers (e.g. VPN-, P2P-based) with new challenges and opportunities for network management.

In the framework of the EU funded 4WARD project we are currently studying new management concepts based on autonomous, self-organising entities with more inherent management capabilities (INM) in the network nodes. In this way, the reliability and scalability of centralized management can be improved. Completely distributed management approaches are investigated, e.g. for dynamic network environments where centralized approaches fail. In this paper, we study the INM business value in a qualitative analysis by describing scenarios where INM is expected to improve performance and save costs.

KEYWORDS

Autonomous, self-organizing network management, OpEx, CapEx, business value

1 INTRODUCTION: MANAGEMENT OF THE FUTURE INTERNET

Fixed and mobile broadband access is developing towards a standard for business and residential home users with more and more attractive services being integrated on Internet platforms. The corresponding cost and business models cover multiple layers from infrastructure to service provisioning. New Internet services continuously create new revenue opportunities in online markets, while network operators can profit from an increasing subscriber base attracted by a steadily extending service spectrum. They facilitate the launch of new services and in addition help to reduce the effort and costs of service and network provisioning.

The INM paradigm presented by the 4WARD project allows for an integrated approach to service and network management while embedding intelligence into the network and minimizing the requirements for human intervention.

Since a quantitative prediction of the INM business value for developing new services seems speculative, we focus on describing main cost saving effects through INM for operational (OpEx) and capital (CapEx) expenditure. Driving forces of the expected savings and business support can be seen in

- Self-organization and automated processing, reducing the need for manual intervention in current network operation
- Distributed monitoring and control for better situation awareness, resulting in faster and more precise proactive and repair processes
- Virtualization concepts which include autonomous management in a common framework
- Improved control and reporting functions for business management

Savings in CapEx are expected through combined situation awareness and optimization for more efficient network resource usage [2][16]. Since current preferences for over-provisioned networks also contribute to wastage of energy in underutilized electronic equipment, optimized resource management supports CapEx saving together with a trend to green information and communication technology (ICT).

We summarize OpEx and CapEx aspects in network management from a general point of view in section 2, while section 3 addresses main contributions and shifts to be expected from autonomous and self-organising approaches. Profitable scenarios for those approaches with focus on dynamic networking and overlays are discussed in section 4 followed by the conclusions.

2 RELEVANT CATEGORIES IN OPEX AND CAPEX

Operational (OpEx) and capital (CapEx) expenditures are interconnected issues [16]. It depends on the network environment whether OpEx or CapEx represents the major portion of the overall costs of the network and service provisioning [3][5][16]. In a Future Internet environment, automated technologies and functions to facilitate management tasks may initially shift costs from OpEx to CapEx but in the long term will reduce the total costs as simplifying standard solutions. The level of achievable self-organization is different in fixed backbone and access network types, as compared to wireless and mobile networks.

Network layers are another criterion for differentiation. Virtualization concepts can provide a completely self-organizing environment for some task on top of a physical network infrastructure. Developments towards virtual concepts can already be seen in application layer overlays, e.g. peer-to-peer networking, whose economic aspects are the main topic of the EU project SmoothIT <www.smoothit.org>.

In this article, we focus on processes in operational networks. Therefore first installation costs in OpEx as well as CapEx for setting up a network, buying equipment etc. corresponding to processes defined in [14] are not considered. We assume that the same categories of operational processes will be as relevant in Future Internet as they are in today's networks. However, the way these processes are designed and implemented will make the difference in a Future Internet scenario. The costs occurring in an operational network can be divided into the following categories related to main management tasks and processes:

1. Continuation of normal network operation
   The cost to keep the network running in a failure free situation including space, power supply, leasing equipment, e.g. fiber rental etc.

2. Maintenance and monitoring
   The cost to maintain the network and to operate the network with awareness of failure events based on monitoring of the network and its services

3. Failure handling and recovery
   Failures in the network have to be repaired on a case by case basis triggered by alarms, if this cannot happen in routine operation. Reparation may lead to actual service interrupts depending on the protection scheme.

4. Planning and updating of an operational network
   This category includes all planning performed in an existing network which is up and running, resulting in long term upgrade processes for increasing traffic and resource demands as well as short term workarounds and (re-)optimization, upgrades or replacement of outdated software and hardware components.

5. Service management and provisioning
   Processes set up to provide and control previously negotiated services to customers, usually defined via service level agreements (SLA)

6. Business management, marketing, pricing, sales
   Business and service managers govern the network to support a service portfolio through a set of business decision processes. They use policies at the top level when developing new end-user services.

Network management has direct influence on the first four categories and interacts with service and business management through predefined interfaces and processes. Studies on the distribution of OpEx for network operators [10][12][13] [15] attribute roughly 27% to marketing and sales, 24% to customer and IT support services, 22% to network elements and another 27% to interconnection and roaming.

Table 1 shows that the latter portion has higher estimates for mobile and lower for fixed operators. Interconnection costs also differ with the size of network and Internet service providers (ISPs). Globally operating tier-1 ISPs often provide Internet connectivity for smaller ISPs but can profit from peering contracts with them.

Table 1: OpEx distribution estimates

OpEx Category	Estimate for mobile operators (GSM) [10]	Estimate for mobile operators [13]	Estimate for fixed operators [13]
Network Elements	20%	20%	25%
Marketing & Sales	26%	26%	30%
Customer Service	15%	8%	10%
IT Support & Service	13%	11%	15%
Interconnection, Roaming	26%	35%	20%

3 THE IMPACT OF INHERENT NETWORK (INM) MANAGENT ON BUSINESS VALUE

Basic concepts of the INM approach are decentralization, self-organization and embedding of functionality and autonomy, as illustrated in Figure 1. This enables management tasks or subtasks to be delegated from special management platforms to a self-organizing management plane inside the network. At least an appropriately selected part of management functions can be perform inherent to the network, for instance, reconfiguration or self-healing in an autonomic manner.

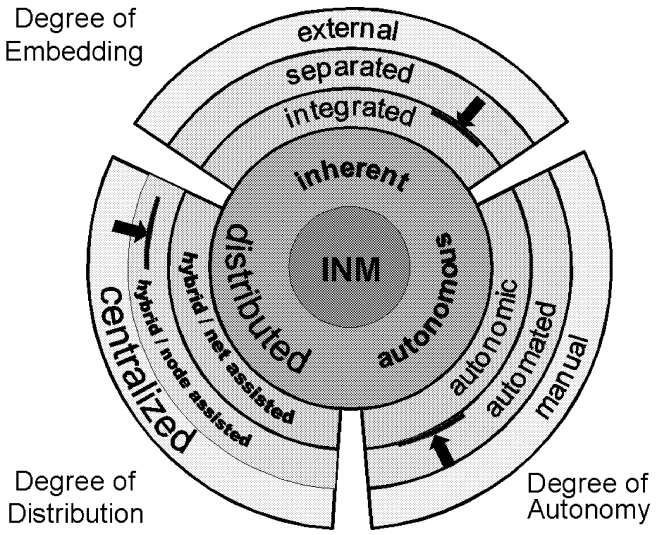


Figure 1: Developments towards INM

Real-time monitoring or, more general, situation awareness is important for several reasons. First, monitoring, i.e., the process of acquiring state information from the network and its nodes, is a fundamental precondition to operate a system efficiently. Monitoring functionality is generic since it can directly support the management of the underlying concepts being developed within EU project 4WARD [2], i.e., network virtualization, generic paths and network of information.

3.1 OpEx reduction through self-managing instances

In nowadays heterogeneous network environment, different domains establish their own degree of embedding, autonomy and abstraction, depending on the technology, applications and administrative goals. Wherever a higher autonomy level or more flexible adaptation is achieved, expenses for manual intervention are reduced with consequences for OpEx [1][10].

Centralized management in large scale networks is often subject to uncontrolled floods of alarms being forwarded to the network operation center in failure cases. Self-managing entities in INM are expected to enforce efficient local control loops which can filter and forward more consolidated reports instead of spontaneous data records. INM can aggregate control messages within entire sites and areas. It will not be necessary to have every network element connected to a central network management system, which may detract from scalability on large platforms. While this can reduce overhead and facilitate failure analysis in large scale networks, it is a precondition for deploying small home or building networks, which have to work in an autonomic manner without an operator being involved and otherwise have to generate failure reports to be understood by untrained people.

3.2 CapEx reduction through situation awareness and fast adaptability

Giving preference to a higher autonomy level, self-organizing schemes can also profit from ICT technology trends towards steadily improving performance. As a side effect of increasing bandwidth, monitoring and control cycles can be executed at higher frequency, resulting in immediate and more precise situation awareness for distributed network functions [11]. Consequently, resource optimization, load balancing and rerouting can adapt to traffic shifts and prospective failure scenarios in shorter response time, which makes them more reliable and robust.

As an example, we consider traffic engineering on broadband access platforms, which has to cope with link upgrading processes for increasing Internet traffic and has to take care of failure situations by providing backup paths [4][7].

(G-)MPLS or Ethernet with multiple spanning trees are technologies that allow to establish load-balancing traffic paths. They can be pre-computed, not only for the current topology and traffic load, but also for modified topologies to upgrade link bandwidth and to include relevant failure scenarios. In the latter case it is still time critical to reconfigure the paths to circumvent a failed link in an operational network. Therefore additional capacity is usually provisioned to overcome non-optimized intermediate stages after topology changes due to failures or upgrades.

INM functionalities for faster situation awareness and adaptability through automated online processes can reduce those time gaps subject to unbalanced or instable load in the network and, as a consequence, allow for reduced over-provisioning for such situations. Experience with traffic engineering in IP backbones [7] has shown that a 20%-30% increase in utilization – or a corresponding decrease in provisioned bandwidth of IP routers and transmission equipment – is possible when the network can be kept in optimized load-balanced regime by fast redirection of transport paths from overloaded resources.

3.3 Support for business and service management

Business and service management relies on and interacts with network management, e.g. for evaluation of the network status and performance over time. Advanced INM monitoring and automated reporting again improve the timeliness and precision of evaluated data and support SLA fulfillment etc.

As a long-term vision, self-adaptation and autonomous management may vertically integrate network resource monitoring, planning, administrative policies with business and market strategies. For example, pricing for access and IP services may be made dependent on temporary bottlenecks and restricted scalability of network resources. In the opposite way, marketing campaigns and new pricing for web services may automatically trigger support actions in resource provisioning and management for prospective shifts in demand.

4 Scenarios and case studies

4.1 The need for inherent network management in dynamic networks

Dynamic network environments represent an area where classical management and routing concepts fail and a clean slate approach seems indispensable [9]. The IETF has set up a working group on ROuting over Low power and Lossy networks (ROLL, www.ietf.org/html.charters/roll-charter.html) to study whether existing protocols can be adapted or what new methods have to be set up for networks with unreliable nodes and links.

Network management has to be included as a next step [9] [17]. Therefore concepts of inherent network management (INM) are relevant to achieve fast local situation awareness and control, which can be exchanged and evaluated in larger domains via distributed and gossip-based algorithms [18], random walks [6] or reinforced learning [9]. The scope of dynamic networks covers heterogeneous environments e.g. for home, building and urban networks or communication for recovery from catastrophes and military applications. When new concepts can be realized to manage and operate networks with a high level of dynamics, they should be flexible to handle less dynamic scenarios and to support a centralized view of the network as far as possible.

INM may contribute in a long term perspective to a generalized network management framework spanning a wide range from distributed architectures for dynamic systems to centralized schemes for stable topologies [8][9]. In this way migration steps towards self-organizing and autonomous components are also prepared for classical management areas.

4.2 INM for Femto-cell networks

The growing field of femto networks presents an interesting area to which the INM principles lend themselves or could at least be studied within. The way they are designed, in combination with the femto characteristics make them a promising candidate for evaluating INM designs. Since femto cells combine fixed and mobile networks, they are right at the heart of convergence. A usual scenario is shown in Figure 2. Additionally, they implement a potentially vast area of the Internet's network edge, including end user's homes.

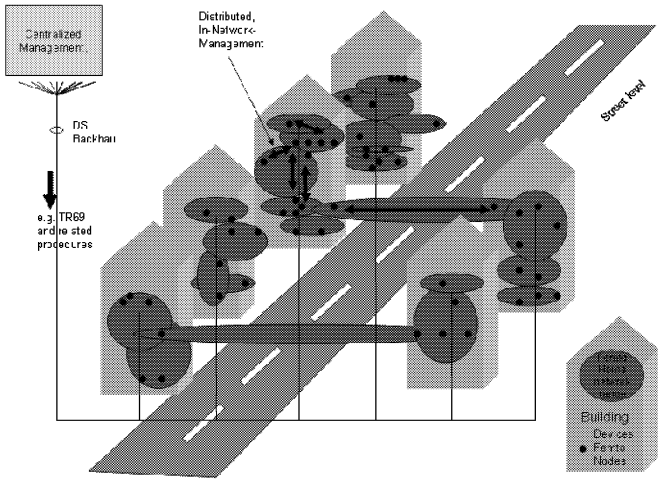


Figure 2: Femto-cell network example with relation to INM

What makes them even more special is the fact that they are designed in a way that requires a comparatively very low administrative, management and also planning effort – and this is part of the business case behind, which, of course, mostly builds on spectrum economics. However, there is a potential for improvement, which is two-fold:

- Femto-designs often use a so-called femto-access-gateway, a centralized network element coordinating network access for femto installations. We assume that parts of this functionality could be executed "inside" the femto-cloud such that less capacity is needed on the gateway.
- Since there is almost no spectrum management for femtos, there are obvious options for improving, for example, anti-interference management, which would be implemented according to the INM principles, i.e. mainly by letting femto nodes communicate and negotiate with each other. While currently TR-69-like standards describe the centralized management, this could be complemented by an INM inter-femto signaling and management protocol.

All in all, we believe that femto-cell networks, in particular the upcoming, LTE (flat architecture) based femtos, are good candidates for evaluating INM concepts. The business value can be modeled according to both presented improvement options. If the INM principles can be validated and assured in combination with cost or capital expenditure savings for the femto application, it should be evaluated for other scenarios, too, such as MPLS based core networks, for example.

4.3 Inaccuracy in situation awareness: A standard OLSR case study in MANETs

In a case study we focus on using the optimized link state routing (OLSR, RFC 3626, <www.ietf.org/rfc/rfc3626.txt>) as the standard routing protocol of the IETF MANET working group to be extended for situation awareness [8]. We consider wireless broadcasting nodes placed at different locations spread over an area, initially without mobility. We assume that two nodes can directly exchange messages when their distance does not exceed a common transmission range. In this way, the transmission links between nodes and the network topology is determined by the node locations.

The nodes route messages via OLSR over multiple hops, including Hello and topology control (TC) messages at default intervals of 2 and 5 seconds, respectively. We consider the queue size of messages that arrived at a node but are not yet forwarded towards the destination as a quality of service (QoS) measure. OLSR is extended to distribute the queue size such that each node is aware of the queue size at all other nodes subject to a delay until routing updates are received between nodes. The queue size information is not used for load balancing. To store the QoS-related state associated with a node, a new field is added to the neighborhood information base and to the topology information base maintained by the protocol. To populate these fields, the message format of Hello and TC messages were extended as well. Table 2 summarizes the parameters of the modeling and simulation parameters.

Table 2: Simulation Parameters

Simulator Parameters	
Network Type	IEEE 802.11
Propagation model	Two-Ray-Ground
Mobility model	Static
Transmission range	250 m
Network topology	50 nodes randomly located in 1 km^2
Traffic model	20 random source-destination pairs, constant messaging intervals: 0.2s, 0.14s, 0.09s, 0.04s, 0.02s for different traffic load levels
Packet size	128 Byte
Queue	for max. 50 packets; Tail-Drop
Simulation time	200s; 50s start phase not evaluated

In this scenario, we study the absolute difference between the current queue size at a node and the aged information about it, which is available at the other nodes through routing messages at the same time, as a measure for inaccuracy of the routing information state.

Figure 3 combines the evaluation of deviations in queue size information with the age of the information at the nodes. The *k*-th column gives the mean deviation for nodes whose information is aged between *k*-1 and *k* seconds or is larger than 21s in the last column.

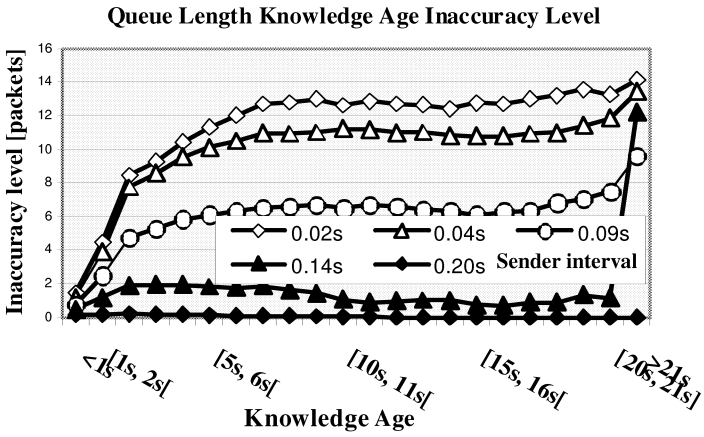


Figure 3: Inaccuracy in routing data and situation awareness

As can be expected, the deviation is essentially increasing with the traffic load, which varies for the five curves by a factor of 10 from low and medium load up to congestion. Inaccuracy is also increasing with knowledge age, although not monotonically.

We also analyzed the impact of sending more frequent Hello and TC messages but did not obtain a positive impact on the overall inaccuracy level. In order to reduce knowledge age, we enhanced the routing by a probing scheme. The current state of a node is looked up by a probing message which is triggered by a threshold for the age of information. Probe messages are fully exploited by updating the new status information for all nodes on their paths as well as for nodes which receive the information due to broadcasting in the surrounding of the path.

The probing scheme was able to halve the mean knowledge age from about 10s to 5s, but in spite of the improvement in knowledge age, probing was unable to reduce the inaccuracy in the queue size estimates. Again, this can be explained from the results in Figure 3, since the inaccuracy level stays almost constant for knowledge ages of 5 and more. Only for ages below 3 seconds essential improvements of the accuracy are observed.

This study for fixed randomly chosen topologies shows that the overall inaccuracy of routing and management information can be high in wireless multi-hop communication. In addition, mobility or high churn will affect reliable transport of routing messages and will increase inaccuracy due to route flapping and loops.

When high dynamics impedes convergence of link state and distance vector routing protocols, then flooding and random walks can be used as well as methods combining flooding, random walks with partial routing knowledge [6].

CONCLUSIONS

Although centralized management approaches are expected to persist in fixed broadband access platforms in near future, autonomic and decentralized approaches will develop in dynamic network domains which currently are becoming more important in several areas including MANETs, sensor networks, overlays and virtual network environments.

If inherent management solutions can be successfully established in dynamic networks, where a central view isn't applicable then self-organizing entities and functions can be added bottom-up to areas under central control. In this way, central management is not expected to be replaced, but to be enhanced by local control, filtering and aggregation of monitoring data and alarms for improved reliability and scalability.

The implementation and evaluation of further scenarios requiring INM concepts in the future Internet is currently under study by the EU 4WARD project team.

ACKNOWLEDGEMENT

This work was developed as a part of the work package on InNetworkManagement (INM) of the 4WARD project. We would like to thank the project partners for providing a framework for business value evaluations of this paper. Especial gratitude we owe to Catalin Meirosu, Ericsson, for fruitful discussion and valuable hints while preparing the paper.

REFERENCES

[1] 3rd Generation Partnership Project, Technical Specification Group TSG RAN, Evolved universal terrestrial radio access network (E-UTRAN): Self-configuring and -optimizing network use cases and solutions, 3GPP TR 32.816, Release 8 (2008)

[2] 4WARD Deliverable D-4.2: In-Network Management (INM) Concept, G. Nunzi, D. Dudkowski (eds), FP7-ICT-2007-1-216041-4WARD/D-4.2, <www.4ward-project.eu>

[3] M. Amirijoo et al., Use cases, requirements and assessment criteria for future self-organising radio access networks, Proc. International Workshop on self-organizing systems (IWSOS), Vienna, Austria, (2008) <www.fp7-socrates.org/ ?q=node/10>

[4] R. Cohen and G. Nakibly, Maximizing Restorable Throughput in MPLS Networks, IEEE Infocom (2008)

[5] A. Ferreiro et al., Migration guidelines with economic assessment and new business opportunities generated by NOBEL phase 2, IST IP NOBEL, deliverable D 2.4 (2008) 1-208 <www.ist-nobel.org/Nobel2/imatges/D2.4_final%20version.pdf>

[6] G. Haßlinger and T. Kunz, Efficiency of search methods in dynamic wireless networks, Proc. Wireless and Mobility, Barcelona, Spain, Springer LNCS 5122 (2008) 142-156

[7] G. Haßlinger, S. Schnitter and M. Franzke, The efficiency of traffic engineering with regard to failure resilience, Telecommunication Systems, Vol. 29, Springer (2005) 109-130

[8] T. Kunz, On the inadequacy of MANET routing to efficiently use the wireless capacity, Proc. IEEE Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob 2005), Montreal, Canada (2005) 109-116

[9] M. Lee, D. Marconett, X. Ye and S.J. Yoo, Cognitive network management with reinforcement learning for wireless mesh networks, Proc. IPOM workshop, ManWeek (2007) <sierra.ece.ucdavis.edu/documents/2007/IPOM2007_BenYoo_ 32795.pdf>

[10] M. Lähteenoja and B. Olsen (Eds.) Techno-economics of integrated communication systems and services: OPEX models, Celtic project ECOSYS, Deliverable 6 (2005) <www.celtic-ecosys.org>

[11] A. Manzalini, F. Zambonelli, Towards autonomic and situation-aware communication services: the CASCADAS vision, Proc. IEEE Workshop on Distributed Intelligent Systems (2006) 383-388 <polaris.ing.unimo.it/Zambonelli/PDF/IEEE-DIS.pdf>

[12] NGMN Alliance, Informative list on SON use cases, Annex A of use cases related to self organizing networks: Overall Description, Version 1.53, Ed. F. Lehser, Next Generation Mobile Networks Ltd. (2008) 1-35 <www.ngmn.org/uploads/media/NGMN_Use_Cases_Self_Organising_Network_2_02.pdf>

[13] T. Rokkas, D. Katsianis, D. Varoutas and T. Sphicopoulos, Fixed mobile convergence for an integrated operator: A techno-economic study, Proc. 18. IEEE Symopsium on personal, indoor and mobile radio communications (2007)

[14] TeleManagement Forum, Enhanced Telecom Operations Map® (eTOM) The business process framework for the information and communications services industry, GB921 Vers. 4.0 (2004) <www.tmforum.org> corresponding ITU-T standard M.3050

[15] S. Verbrugge et al., Operational expenditures for telecom operators, Proc. 9th IEEE Conference on optical network design and modelling (2005) 455-466

[16] S. Verbrugge et al., Methodology and input availability parameters for calculating OpEx and CapEx costs for realistic network scenarios, Journal of optical networking Vol. 5 (2006) 509-520

[17] S. Wallin and V. Leijon, Rethinking network management solutions, IEEE IT Pro (Nov./Dec. 2006) 19-23

[18] F. Wuhib, M. Dam and R. Stadler, Gossiping for Threshold Detection, accepted for 11th IFIP/IEEE Internat. Symposium on Integrated Network Mgnt., New York, USA, (June 2009)

EGOVERNMENT APPLIANCE PROBLEMS ON FIRST DEGREE SELF GOVERNMENT OF GREECE

Dimitrios S. Goulas,
MBA student,
Faculty of Social Sciences,
Hellenic Open University
Patras, Greece
E-mail: dgoulas@patras.gr

Georgia N. Kontogeorga
Candidate doctor
Department of Business Administration
University of Patras
Rio-Patras, Greece
E-mail : kont_georgia@yahoo.gr

KEYWORDS
eGovernment, Local Self-Government

ABSTRACT

Greece, in the frame of appliance of eGovernment, the last years has made significant steps which have changed the way its public services work.

We made a case study of the municipalities of Western Greece in which we examined the level of growth of eGovernment as well as the degree of exploitation of chances given by the various operational programs of the European Union, apart from their own budget, in order to ensure the essential infrastructures and to go towards a higher level of eGovernment for the Greek municipalities.

In this project we locate and analyze the problems that exist in the exploitation of ICT and subsidies from EU, we analyze the reasons that cause them and we search and suggest ways for improvement for the future, in view of the implementation of NSRF.

INTRODUCTION

eGovernment is defined by the EU organisation as the use of information and communication technologies in public administrations combined with organisational change and new skills in order to improve public services and democratic processes and strengthen support to public policies. The Greek municipalities today have the ability to take advantage of the European Union programs and subsidies to achieve a better level of eGovernment, with proper substructures – organizational and technological-, and advanced services. Having already finished the 3rd CSF (Community Support Framework - KPS) and starting participating in NSRF (National Strategic Reference Framework –ESPA), the results show that the municipalities have problems in adapting and assimilating new organizational methods and new technologies. Moreover the available EU subsidies aren't being absorbed and we have a huge waste of money without the expected results.

This project aims to look closer the situation in Greek municipalities regarding eGovernment and track down the existing problems and reasons, in order to propose solutions for a better utilization and a more efficient eGovernment.

We couldn't make the research on data of all Greek Municipalities so we followed the case study method for the region of Western Greece. Even in that case we used layered sampling of the total municipalities according the population. We examined the eGovernment growth level and the utilization grade of the funding chances of the EU's operational programs and in this way we spotted and analysed the existing problems in the utilization of the ICT and the EU fundings for eGovernment.

For implementing this project we used, apart from the existing bibliography about eGovernment and the information from internet about the absorbency and the funding claiming, diverse interviews and a questionnaire applied on the municipalities about internet and internet services' use, the use of software applications, the existence of web site/portal and web services, the existence of organizational departments and employees relevant with informatics, networks, EU programs etc.

We start our project by concentrating on the region of Western Greece and examining the ICT use and the participation range of the municipalities of this region in EU programs and actions on eGovernment. On this we spot the affronted problems in funding claiming, implementing and utilizing EU programs in order to have better and suitable adoption of ICT in their services to the citizens and their employees work.

According the above we make suggestions to overcome the existing problems and obstacles the municipalities affront so that they can better utilize the EU programs and improve and develop the use of ICT and eGovernment in general in the NSRF. Finally we result in several conclusions giving ideas about how municipalities can work in the future for a more effective eGovernment.

THE REGION OF WESTERN GREECE

As a result of the recently reform under the Kapodistrias Municipal Code (1995) which came into force in 1999 Greece consists from:

- 13 regions (peripheries)

- 51 prefectural authorities, including 3 extended prefectural authorities and
- 1.031 municipalities (130 urban municipalities – *dimi* and 901 rural communities – *kinotites*)

The region of Western Greece consists of three Prefectures, namely Achaia**,** Ileia and Aitoloakarnania**.** Following the recent administrative division (Law 2539/97), they are further divided into 74 OTA (Local Government Organizations), 72 Municipalities and 2 Communities, as indicated in the table 1 below (www.ditikiellada.gov.gr):

Table 1: Structure of the Region of Western Greece

	AITOLOAKARNANIA	ACHAIA	ILEIA	TOTAL
MUNICIPALITIES	29	21	22	72
COMMUNITIES	--	2	--	2
TOTAL	29	23	22	74

The Municipality of Patras is the most heavily populated (167.602 according to the National Statistical Agency's 2001 census). However, there are another sixteen (16) municipalities in the Region with a population of over 10.000 inhabitants (Agrinio, Messolonghi, Nafpaktos, Amfilochia, Oiniades, Aigio, Rio, Amaliada, Ancient Olympia, Zacharos, Pirgos, Skyllountos, Messatida, Dimi, Gastouni and Vouprasia). A further twenty-seven (27) municipalities have a population of between 5.000 and 10.000 inhabitants (2001 census).

Western Greece's population is 741.282 people, making it the fourth largest (in population) region in Greece, accounting for 6,8 % of the country total. About 45,1 % of the population is urban, 11,4 % is semi-rural and 43,5 % is rural. It covers an area of 11.350 km^2, which is about the 8,6 % of the total country area.

COLLECTING DATA

To carry out our study we drew up a questionnaire which was sent to diverse municipalities of Western Greece, to find out their degree of involvement in eGovernment initiatives. We obtained responses from 19 local governments and by using layered sampling according the population we got the information we wanted, relevant with the ICT use and eGovernment appliance at the municipalities of Western Greece. The information we collected concerns the period till January 2009.

To have a better view of the existing conditions we took some interviews from the employees who filled in the questionnaires and were responsible for ICT and Infosoc projects and from elected representatives with relevant responsibilities of the municipalities' councils. Very helpful was our personal experience and by using internet information we found out more specific information about some of the implemented or under implementation EU programs.

Our research was splitted to the following parts: the organization of the municipalities, the ICT level they are, the eGovernment steps they have done and information about their participation to EU funded programs.

DETECTED PROBLEMS IN ADOPTING ICT AND E-GOV

There is a lack of high educated personnel on informatics in most municipalities (62% had none specialized employee and the rest 38% need more). Municipalities need specialized people so that to support and promote the use of ICT, people that will understand the needs and will plan the proper infrastructures for the adoption of the new technologies. That is one of the most important reasons that municipalities have fallen behind in progress and ICT evolution.

Moreover, lack of training of employees causes scarcity in simple users within public agencies with ability to operate the Information Systems used. In addition to this, the problem gets bigger as the employees are not interested in new technologies as they have learnt to work in a specific way and they don't want changes. The administrations haven't changed the way employees think and perceive the ICT and the facilitations they provide.

We must mention the fact that most elected representatives that cover administrative posts at the municipalities are not very familiar with the new technologies. This entails lack of vision that could boost ICT use and help the Greek government to implement its aim for eGovernment.

There is also a lack of proper organizational structures for the support of ICT. 74% of the municipalities have no office, department or directory relevant with new technologies and even the most populated are keeping to be supported from only one office or department with minimum number of employees. So, the organizations cannot define and assign responsibilities relevant with ICT to departments although almost all of them use information systems and technologies. Their support is depended on private companies that support them or the good will of their employees -that even in this case cannot program the needs and the procurements of the proper hardware and software-.

Related with all the above is the existence of insufficient technical infrastructures. If we add the fact that the investments in ICT had been rather low the result is that, in favour of the temporary confrontation of the municipalities' needs, we have problematic substructures for networks, information systems and hardware systems, which cannot get used as base for new innovative movements for appliance of new technologies without serious and costly changes.

Although all municipalities have -or would like to have- applications for logistics, registries, management of municipality police or immigrants' applications etc, they use different applications and Information Systems which are developed separately for each individual organization to meet their differentiated needs. They are not considered as parts of a unique and global informative infrastructure for the Public Administration as a whole. So, as far as knowledge management is concerned, the fact that most existing Information Systems are not integrated but they operate independently from each other makes it difficult to ensure efficient information diffusion between administrative

functions and effective delivery to the citizens. So, many times we have conflicted or repeated and wasteful data.

Finally, even though 68% of the municipalities are implementing portals and give the ability to citizens for searching personalised information, certificate issuing and making some transactions, they use different framework for the design of the websites and they don't follow a common design template for their websites.

DETECTED PROBLEMS IN CHALLENGING SUBSIDIES AND IMPLEMENTING EU'S PROGRAMS FOR E-GOV

The most essential barrier for the implementation of EU's programs is the inefficient organizations, which didn't predict the needs for managing EU projects. So, municipalities hadn't the appropriate personnel for these tasks and even if they had they couldn't utilize them because they hadn't suitable structures. The majority (84%) had no structure and the rest 16% had substandard structures. This problem was even bigger because of the fact the procedures are new and the training in this field is inexistent or narrow.

The bureaucratic organization of public services has also significant role. The existence of many hierarchical layers involved in the public business processing impedes decision making and assignment of duties. So, except delays, which in many cases are crucial for the projects, it demands knowledge on how other services work and how they should cooperate, which not exist.

Moreover the Greek ministries, being the main subsidies' administration backbone, had their own problems which made more intense the problems the municipalities had. The responsible personnel for the programs at the municipalities didn't have the guidance and help they needed on the new for them procedures and this made their work more difficult.

Till recently all municipalities had no strategy in the field of eGovernment and only lately 68% of them conducted Business Plan. Having no strategy, Municipalities just wanted –and that's the only thing they partially achieved- to absorb European Union's subsidies on new technologies, without managing to take advantage of them to build well-designed, modern and effective organizations. It is obvious that the national strategy didn't "pass" from the ministries to the local governments and it hadn't the results it should. The local authorities didn't understand the importance of setting objectives and of well-organized subsidy claiming in having progress and becoming effective and operational in a progressive and always transforming society with the new technologies playing the leading role.

Related to this is that municipalities hadn't completed and documented studies for actions they wanted to take and prefigures, including budgets and prerequisites, so that to have the ability through European Union's programs to modulate them and get funded for their implementation. The municipalities till now were taking to make offhanded and segmental studies to catch up the deadlines for the proposals. The result was extra problems to the implementation of the projects because of the inadequate infrastructures and the absence of "inside" help from people who know the subject and are trained on it.

The above lack and the fact that local governments had available delayed and not thorough information on Information Society's projects had as result the failing or delaying to deposit proposals and simultaneous had not taken them into account at the designing of local projects, with which they could cooperate or being complement.

Finally, there is no specific algorithm for the procedures and municipalities do not follow standards for composing studies, proposals' submission, monitoring and implementation procedures, quality and utilization of the projects etc, which results in having problematic and uncompleted proposals or segmental implementation or low efficient utilization

PROPOSALS ON ORGANIZATION

Having detected the above problems, Municipalities need to reform their Internal Organisation and the organisations of the Institutions they supervise and include special departments for writing studies/proposals, proclamations, supervising and utilization of projects, so that to obtain flexible and modern forms of management, implementation and utilisation of ICT projects.

In this way they can promote their certified administrative sufficiency for ICT projects according to quality standards or special national or international standards. Besides, for their participation to NSRF and funding claiming from EU they need to be certified according the Greek Standard of Administration and Project Management that is prepared by ELOT and is almost ready.

There should be put in charge technical support executives in each project, different for each prefecture or region that will be responsible for the coordination, the studies etc for each region. The technical support to the beneficiaries is essential, because of the technical nature of the projects (for which the municipalities usually do not have the necessary know-how) and also they don't know how to confront with the utilization of the implemented infrastructures.

Municipalities must define their Strategic plan regarding growth of ICT use and eGovernment. It is important to know our aims and targets in order to search suitable methods and means to succeed these. Knowing the Operational Programs included in NSRF and the targets EU has, municipalities can harmonize their strategies and plans in order to achieve the necessary funding for costly projects.

It is also necessary municipalities to define an enterprise system with indicators of measurement and evaluation of the appliance degree of their strategy and the achievement of their objectives. With a tool like this, they could periodically locate the problems and tendencies in order to take suitable correction actions to achieve their aims. EU has already introduced the Common Assessment Frame (CAF) which municipalities can apply in order to see the quality level of their services. Besides the Greek Standard of Administration and Project Management which will be applied for the

administrative sufficiency for NSRF they can also do a step forward applying for ISO 9001/2000 and adopting Total Quality Administration.

Moreover, if the organization level comes to a better level it would be useful to prepare projects' proposals - actions for the 4th Programmatic period. Technical services have already an idea for the EU's projects and the procedures for their implementation. Instead of waiting for the Calls so that to take the proper actions, they can start preparing the necessary studies, following the directions of their Operational Strategies and utilising the Operational Programs of NSRF.

One more suggestion is that there could operate a support Office for the implementation and operation of projects from municipalities that could develop a guide with the experiences (positive and negative) and lessons learned of the funded projects that the municipalities have already carried out (benchmarking) and provide the following services:

- Support via telephone call, fax or email
- Supervising of implementation and operation of the projects
- Portal for information and support for users in the below subjects, at least, with regard to the implementation of projects of information society (InfoSoc).

PROPOSALS ON EDUCATION/TRAINING

Each Municipality and community must take care to have at least two executives at their staff with training on ICT and train their executives in the planning of projects and Internet use, taking care for continuous promotion of modernisation of local self-government, having as axis the new technologies. They can utilize e-learning infrastructure of SYZEFXIS for the training in the ICT of their executives.

To the same direction Municipalities should create executives that can also deal with InfoSoc projects. This means Municipalities should take trend for special training of executives, setting direction (depending on their size) for their projects administration, the frame for wider collaborations of municipalities with the private sector, the exploitation and the transmission of know-how internally etc.

It is also necessary to implement training seminars for the familiarization of all employees with ICT especially on basic office automation programs and the use of internet and electronic mail. The adoption of an organizational information system produces radical changes at the procedural and work practice level. As a result, new organizational models need to be introduced as well, focusing on knowledge management and extensive training.

Municipalities must try to further grow their eGovernment services' supply and promote electronic services to the citizens trying continuously to cover their, always growing, demands. Through e-learning, special seminars organized for citizens and advertisement of the municipalities' electronic services it could be much easier.

Last of all, through the Ministry of Interior, could be composed and diffused toolkits for support of programming, planning, implementation and utilization of the projects. These efforts are supplemented or they are identified with a complete pack of directives (step to step) that is addressed to all Local Self-government and can solve various subjects (i.e. preparation, implementation and operation of the projects).

PROPOSALS ON COLLABORATION

It is important municipalities to improve or create InterMunicipality enterprises that can take the responsibility to conduct studies for the beneficiary municipalities, help them supervise the implementation of projects and make proposals or even undertake their utilization. These enterprises are more flexible to employ the suitable personnel and to get funded for these tasks.

The InterMunicipality collaborations can take the form of conducting common proposals for NSRF. Adjacent municipalities could implement projects that will benefit all of them and will help for the growth and development of the region. Especially small municipalities, which have not enough personnel or the know-how, should help each other and get all the help the can from bigger and more developed municipalities following the good examples, improving their claiming and implementation abilities.

One more step municipalities should take is the creation of a collaboration frame between the institutions of the public and private sector for the effective implementation of projects of Information Society (Public and Private Sector Partnerships –SDIT). The fact that there is no money for public investments and it is much time-consuming and complicated to organise public institutions makes the collaboration with the private sector seem the best solution, especially for costly and highly stern projects. However municipalities need to be cautious and well-prepared for this new scheme of collaboration.

Last, we must point out that administrations should try to respond immediately to every call or need of the ministries for cooperation and minimise the delay and non-responding phenomena that observed in cases such as the HERMES portal, the National Registry, etc.

CONCLUSIONS

Funding the eGovernment infrastructure and development is quite a challenge for local authorities in Greece. EU's funded programs were a big chance for them to found the assets for important projects and get helped to implement the proper infrastructure and to step forward to its digital future, to eGovernment.

With the completion of CSF III and in the beginning of NSRF, the final result in eGovernment is not satisfactory, with many calls of EU's Operational Programs finding no response, many delays and problems to the implementation of the projects and bad utilisation of the implemented projects.

The problems of adopting ICT we detected were the lack of high educated personnel on informatics and training, the lack of interest of employees in learning more things on new technologies, the lack of eGovernment vision and ICT knowledge of most elected representatives that cover

administrative posts, the non-proper organizational structures for the support of ICT, the insufficient technical infrastructures and the use of different software applications and not integrated Information Systems that operate independently from each other.

The problems in challenging subsidies and implementing EU's programs we detected were the bureaucratic -with many hierarchical layers- organization of public services, the inefficient organizations, the lack of specific procedures and standards for the procedures regarding projects, the absence of strategy in the field of eGovernment, the lack of completed and documented studies for actions they want to take and prefigures, the delayed and not thorough information on Information Society's projects which means slow take-up rate of contracts, the fact that the national strategy didn't "pass" from the ministries to the local governments, the lack of guidance to the responsible personnel for the programs at the municipalities.

Our proposals to improve the situation and to work the municipalities more organised and prepared for NSRF were categorised to organisational, educational and collaborations.

Regarding organisations Municipalities must reform their Internal Organisation and to include special departments for writing studies/proposals, proclamations, supervising and utilization. There should be put in charge technical support executives for each project, different for each prefecture or region that will be responsible for the coordination, the studies etc for each region. Municipalities should define Strategic plan regarding growth of ICT use and eGovernment and form an enterprise system with indicators of measurement and evaluation of the appliance degree of their strategy and the achievement of their objectives. Preparation of projects' proposals - actions for the 4th Programmatic period is also important. There could also operate a support Office for the implementation and operation of projects so as to work better in the NSRF.

Regarding education/training each Municipality and community must take care to have at least two executives at their staff with training on ICT and that can deal with InfoSoc projects. Training seminars for the familiarization of all employees with ICT are needed as to further grow their eGovernment services' supply and promote electronic services to the citizens trying continuously to cover their, always growing, demands. Through e-training, special seminars organized for citizens and advertisement of the municipalities' electronic services it could be much easier. Central authority should compose and diffuse toolkits for support of programming, planning, implementation and utilization of the projects.

Regarding collaboration It is important municipalities to improve or create InterMunicipality enterprises, InterMunicipality collaborations, Public and Private Sector Partnerships.

FUTURE WORK

What is interesting to see in the future is the measures the municipalities will take in relation to improvements they will have in ICT use and their participation in NSRF. This will show us how these measures contribute to the operability and development of public services and what other measures could be taken. This will help Greece to develop a good level of eGovernment and moreover, will help other new EU members to achieve their digital convergence.

REFERENCES

COMMISSION OF THE EUROPEAN COMMUNITIES (2008) "Preparing Europe's digital future i2010 Mid-Term Review", Volume 3: ICT Country Profiles, Brussels

COMMISSION OF THE EUROPEAN COMMUNITIES (2003), "The Role of eGovernment for Europe's Future", Brussels

European Commission – eGovernment practice (2008), eGovernment Factsheets, "eGovernment in Greece", Edition 10.0

Hahamis P., Iles J. and Healy M. (2005), "e-Government in Greece: Bridging the gap Between Need and Reality", University of Westminster, London UK

Margherita Antona, "eGovernment country report for GREECE", GREECE

Ministry of Economy & Finance/National Coordination Authority (ΕΥΣ/ΥΠΟΙΟ) (2008), MINISTERIAL CIRCULAR "Instructions for Special services (ΕΥΔ) for the confirmation of the administrative sufficiency of beneficiaries 2007-2013 for the period till the beginning of appliance of the Greek Standard of Administration and Management of Projects (Transitional Period)"

Observator for the Information Society (2008), "The use of Information Communication Technologies at the regions and the Local Self-government"

Torres L., Pina V. and Royo S., (2005) "E-government and the transformation of public administrations in EU countries: Beyond NPM or just a second wave of reforms?", Zaragoza (Spain)

DIMITRIOS S. GOULAS was born in Nafpaktos, Greece and went to the University of Patras where he studied Computer Engineering and Informatics and obtained his diploma and MSc. After doing his military service, he worked for a couple of years at the Technical Institution of Messolonghi, the Ministry of Defense and the Municipality of Patras where he is working till now.

GEORGIA N. KONTOGEORGA was born in Patras, Greece and went to the University of Piraeus where she had obtained her diploma in Economics and to the University of Patras where she had obtained Master in Business Administration (MBA). She worked for 8 years at the Minicipalities of Paralia and Patras and now she is working at the Hellenic Court of Audit.

MEDICAL APPLICATIONS

A VIRTUAL ENVIRONMENT TO CREATE SOCIAL SITUATIONS: FIRST STEP TO A VIRTUAL REALITY EXPOSURE THERAPY SYSTEM FOR SOCIAL PHOBIA

Willem-Paul Brinkman
Fatma Inan
Charles A.P.G. van der Mast
Delft University of Technology
Mekelweg 4, 2628 CD Delft
The Netherlands
Email: w.p.brinkman@tudelft.nl

KEYWORDS
Virtual reality, social phobia treatment, presence.

ABSTRACT

This paper describes a study to examine how a social situation can be simulated in a virtual environment, and how to provoke the same behavioral response as in a real life social situation. The aim is to create new Virtual Reality worlds for treating social phobia. Two cases were examined in a virtual environment to assess whether it was able to recreate social behavior of people. First, we examined if participants who enter a crowded room in the virtual environment prefer to take a seat in a chair that is close to them. Secondly, we investigated if participants in the virtual environment have the tendency to physically distance themselves from strangers when choosing a seat in a room full of strangers. We conclude that a social situation can be recreated in a virtual world.

INTRODUCTION

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (1995) social phobia is a persistent fear of one or more social or performance situations in which the person is exposed to unfamiliar people or to possible scrutiny by others. The individual fears that he or she will act in a way that will be humiliating or embarrassing. Traditional exposure therapy (exposure in vivo) typically consists of confronting the feared situation in imagination or in real life (Hodges et al. 2001). Research has proven that Virtual Reality (VR) technology can be successfully implemented in clinical therapy Emmelkamp et al. 2002, Wiersma et al. 2008, Emmelkamp&Powers 2008). Clinicians are using Virtual Reality Exposure Therapy (VRET) to treat anxiety. The exposure therapy takes place in the therapist's office in a computer generated world of the feared situation. The VR gives the therapist greater control over the feared situation, which results in the most effective exposure.
The Delft University of Technology and the University of Amsterdam are collaborating in the frame of the research project "Virtual Reality and Phobias". The Delft University of Technology, section of Human Computer Interaction (HCI), which studies the relation between human and computer. The University of Amsterdam, faculty of Psychology, is concerned with the therapy and psychological aspects of VR.
The Delft University of Technology has developed several virtual environments for treatment of acrophobia, claustrophobia and fear of flying, however a specific virtual environment for social phobia exposure therapy had not been created yet.
The aim of this study is to create a realistic VR social environment that can generate the same response that social phobia patients would experience in the equivalent real life situation. We chose to create a small virtual class room on a campus where people are looking to a film projected on a screen. Not all chairs are occupied. The participants enter the room during the film. This can envoke social phobia. They have to make a choice to sit down on one of the empty chairs. In this paper is described how this virtual lecture room is designed and how a controlled experiment is performed to conclude about the social behavior in this world.

BACKGROUND

Social phobia is not categorized as a specific phobia because, rather than fearing a specific object or situation, people with social phobia fear being judged or embarrassing themselves in front of other people. Social phobia also differs from the specific phobias in that it is more likely to severely disrupt a person's daily life (Kessler 2003).
People with social phobia are well aware of the irrationality of their feelings, but nevertheless feel great apprehension when facing the feared situation. Therefore, they will do anything to escape it and will start avoiding all sorts of social situations, with the result that the avoidance inhibits the person's functional character. In these feared social situations the social phobic will be constantly worried that other people might think they are anxious, weak, silly or strange.
One of the main characteristics of social phobia is the strong anxiety felt before the feared events take place, known as "anticipated anxiety". People start worrying and getting a feeling of fear before confronting the feared situation, once they actually face it, and due to their nervousness, it comes out worse, this leading to an increase in the level of "anticipated anxiety" next time they confront the same situation. A vicious circle is created and which is self generating.

Some people report having had humiliating experiences that triggered their social phobia but others having felt extremely uncomfortable in social situations all their lives. Social phobia often co-occurs with mood disorders, other anxiety disorders and antisocial personality disorder (Neal&Edelmann 2003, Wittchen&Fehm 2003). Once develops, social phobia tends to be a chronic problem if untreated. Most people with a social phobia do not seek treatment for their symptoms (Kessler 2003).
People with social phobia, get more than a little nervous in social situations. Physical symptoms may begin trembling and perspiring, feel confused and dizzy, have heart palpitations, "mind going blank", blushing, stomach ache, and eventually have a full panic attack.
Patients with social phobia have a marked and persistent fear of one or several social or performance situations such as attending social functions, dating, participation in small groups, using a public lavatory or even initiating simple social conversations. As a result, these individuals have great difficulty attending class, working alongside others, eating in public, shopping or even coming to medical attention. They often live alone and work at solitary jobs.
They are sure that others see their nervousness and judge them as inarticulate, weak, stupid or "crazy". People with social phobia may avoid eating or drinking in public, for fear they will make noises when they eat, drop food, or otherwise embarrass themselves. People with social phobia tend to fall into three groups: people with only fear of public speaking; people moderate anxiety about a variety of social situations; people who have severe fear of many social situations, from speaking in public to having a conversation with another person, they all are said to have a generalized type of social phobia.

METHOD

The preference of the participant for certain empty chairs in a room is measured under different circumstances. In order to verify if the prototype virtual environment is recreating the real life situation successfully and is able to provoke the same human response, two test cases were investigated. From Hall (1968) we know that a set of measurable distances exists between people as they interact. He presents public, social, personal and intimate spaces. Studies with standing and seated persons revealed interesting relations between psychological characteristics and interpersonal distance in vivo (Ickinger&Morris, 2001). The following two test cases are assumed to be representative to invoke critical behavior of social phobia patients, and were used for the experiment with the prototype virtual environment.

A. People who enter a crowded room want to sit down as quickly as possible to avoid being the center of attention; they would prefer a chair which is close to them
B. People have the tendency to physically distance themselves from strangers when choosing a seat in a room full of strangers; they would prefer a chair with no one sitting around them

The hypothesis is that people with social phobia, get more than a little nervous in social situations. The majority of non sociophobic people will to a certain extent have experienced a nervous or awkward feeling in particular social situations, but when this feeling becomes extreme and intense, it can be classified as a social phobia. This can be a first step to create a social situation in which social phobia can be controlled by changing parameters (graded exposure).
In this pilot study the participants are for ethical reason not social phobia patients because of the experiment is in a first trial phase. The participants of our experiment are PhD students, staff members or students of Delft University of Technology. For this reason we prefer to choose a situation which can be recognized by the participants.
The social situation we choose for the virtual world is a social event at a university. The participants have to join this social event. The social event is movie night that will take place in a class room of the university building.

The virtual environment scenario

To clarify our design approach we give a broad outline of the situation during the experiment in the virtual environment.
The avatars are sitting on the chair watching a movie. The participant is entering the room too late (which is a feared situation for a social phobia patient), the movie is already playing. They take a look in the room before entering through the window - in order to provoke possibly anticipated anxiety. After standing in the room the participant is going to be navigated through the room, without knowing precisely in which direction. The participants do not have control on the situation after entering the room and cannot get habituated to the situation in the room, since their place in the room is changing continuously.
At certain points, they receive questions about the empty chairs. The participant has to answer the question. The virtual tour ends at the moment after the last question is answered.

The virtual room design of the prototype

Virtual room design has to consist the requirements for provoking anxiety for people with a social phobia. The seating capacity of the room is twenty. There are four rows of five chairs. The rectangular room has a windowed room divider wall. The door of the room is closed (no escape is possible or avoidance of the situation). In front of the room there is a video screen (Figure 1). On the chairs are "living" male avatars sitting, with different clothes (Figure 2). Entering the classroom one sees at the left a room divider wall with a big window. The participant can see already who are in the classroom. The naviation is done automatically by the software as the project leader presses function keys. Also the questions about the preference of chairs to sit in are given automatically by the software with synthetic voice. The classroom and the interaction are modelled with Maya and played interactively with Vizard. The HMD used is an eMagin Z800 3Dvisor.

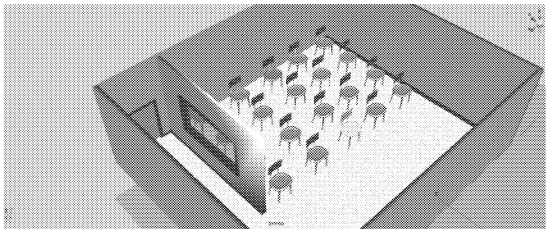


Figure 1: Overview of the virtual classroom.

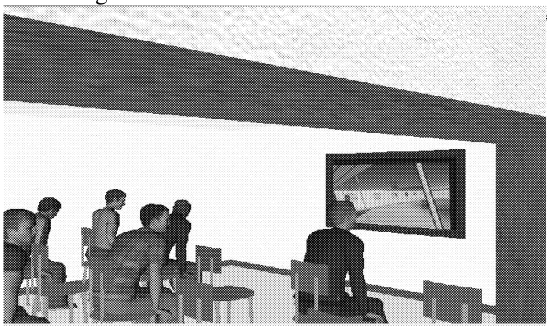


Figure 2: View throught he window in the room divider wall.

THE EXPERIMENT

The animation path starts in front of the room; there are two different animation paths, see Figure 3. The animation paths are chosen randomly. The two different animation paths, which are between-subject variables, are used for different group of participants.
Animation path1 => red chair => yellow chair => green chair => blue chair
Animation path2 => blue chair => green chair => yellow chair => red chair

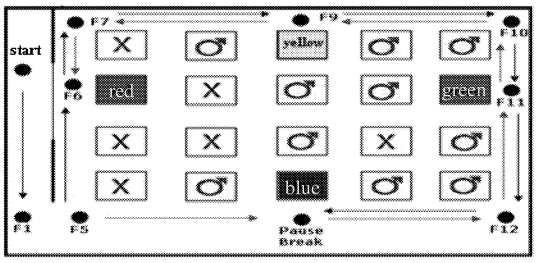


Figure 3: The navigation paths 1 and 2.

We used four measures for our experiment:

1) The Chair preference rate during the experiment with the animation path:
The participants do not get information about the experiment goal before entering the room. They are navigated to certain points in the virtual world and when they arrive nearby a colored chair they receive a question through their HMD a neutral computer generated voice. They have to answer the question: "How willingly would you like to sit on the red/blue/green/yellow chair" and "Please state your answer with the number between 1 and 7" (Likert scale: 1 extreme unlikely, 7 extreme likely).
They are navigated to the next point at the moment they say loudly what their grade is. This repeats for each colored chair till the end of the experiment

2) The preference order of the chairs:
The participant is asked which chair they would prefer first, second, third and fourth if they could walk to a chair when they are standing in front of the group, after navigated tour in the virtual environment.

3) Presence
We used the igroup presence questionnaire (IPQ 2008) to measure the presence. The Igroup Presence Questionnaire (IPQ) is a scale for measuring the sense of presence experienced in a virtual environment (VE).
The scale development process of the current version of the IPQ identified one general item, three subscales. The three subscales can be regarded as fairly independent factors. They are Spatial Presence, Involvement, Experienced Realism. The current IPQ has a total of 14 items.

4) Social phobia
We used the Liebowitz Social Anxiety Scale (LSAS) to assess if our participants do have a social phobia (Heimbert et al. 2006).The more participants with a social phobia, the better we can confirm the effectiveness of our prototype.

The Liebowitz Social Anxiety Scale (LSAS) is a questionnaire by psychologist and researcher, Michael Liebowitz, whose objective is to assess the range of social interaction and performance situations which patients with social anxiety disorder may fear. It is commonly used to study outcomes in clinical trials, see Table 1.

Table 1: The scoring scale of the Liebowitz Social Anxiety Scale (The Anxiety Community 2008)

The scoring scale
55-65 Moderate social phobia
65-80 Marked social phobia
80-95 Severe social phobia
Greater than 95 – Very severe sociale phobia

The scale features 24 items, 13 relating to performance anxiety and 11 concerning social situations. It is not intended for use as a self-reporting diagnosis. The LSAS differs from many of the other measures of social phobia in that it is explicitly situation-based. The participant have to rate his/her fear experienced in a broad array of social situations, as well as to rate the degree to which he/she avoids the situation.
The experiment was performed in the multimedia lab at the EWI-faculty of the Delft University of Technology. The experiment is completed with 22 participants, 5 female and 17 male. The 22 participants were master students (n=15), PhD students (n=5) or staff members (n=2). Some of the participants (n=9) already had virtual reality experience. According to the LSAS scoring scale (Anxiety Community 2008) the participants (19) did not have a social phobia and (3) marked social phobia.

RESULTS

The first step of the analysis was to transform the individuals' chair preference scores (x) into *z*-scores. This would remove individuals' use of the Likert scale. The *z*-scores were based on a participant's average (μ) and standard deviation of the chair (σ) rating.

$$z = \frac{x - \mu}{\sigma}$$

As the red chair and green chair are identical in distance to the film screen, but differ on the number of avatars position around the chair, analyzing preference rating of these chairs would give an insight in the effect of these avatars on participants' preference for these chairs. Next, participants were led pass the chairs in opposite routes, including this information in the analysis gives an insight on participants' preference on preferring a chair rather soon than later when moving through the room. Therefore a MANOVA with repeated measures was conducted. Chair and Animation Path were taken as independent within-subject variables, and preference *z*-score of the two chairs as dependent measure. The analysis found no significant main effect for the chairs ($F(1,20) = 0.13$, $p. = 0.72$) or for the animation route ($F(1,20) = 0.80$, $p. = 0.38$). However, the analysis revealed a significant interaction effect between chairs and the animation route ($F(1,20) = 5.72$, $p. = 0.027$).

Examining Figure 4 with the mean z-score, it seems that the rating of the red chair was higher rated by participants starting with animation path 1: (red-yellow-green-blue),when red chair is asked in the beginning of the path. Participant following the animation path 2 (blue – green – yellow – red), for whom the red chair was the last one on their path, gave a lower score for the red chair. The same was visible for the green chair. The green chair gets a higher score is if the rate for the chair is asked earlier in the path. For animation path 1, the green chair ratings are lower than in case of the animation path 2.

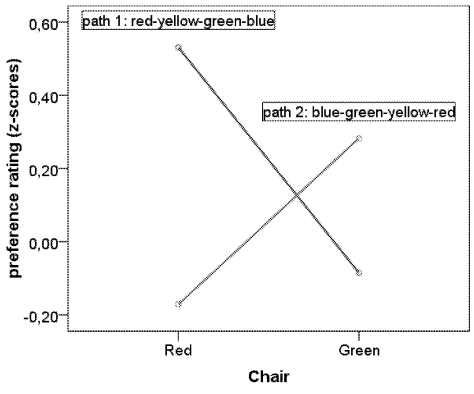


Figure 4: Preference rating chair in relation with the animation paths.

A difference between preferences in chairs was found in the participants preference ordering of the chairs at the end of session. The participants were asked to give a preference order for the colored chairs if they could walk to a chair independent of the animation path.

A Friedman Test on the ranking of all chairs revealed a significant difference (χ^2 (3, N = 22) =12.49, $p. =$ 0.006) between the ranking of the chairs. Furthermore, a separate analysis of the ranking order of the green and red chair with a Wilcoxon Signed Ranks Test also found a significant difference ($z = -2.56$, $p. = 0.011$) in the ranking order of these two chairs. Examining table 6.1 shows that the participants (n=13) prefer the red chair as their first choice. The yellow (n=8) and blue (n=9) are popular as second chair in preference order.

Table 2: The frequency table of participant preference order

preference	Red frequency	Yellow frequency	Green frequency	Blue frequency
1	**13**	**5**	**3**	**1**
2	**4**	**8**	**1**	**9**
3	**2**	**5**	**10**	**5**
4	**3**	**4**	**8**	**7**
Total	22	22	22	22

The LSAS results did not have correlation with any chair preference. The 3 participants who had a social phobia according to the LSAS were not sufficient to make conclusions , still we examined their data and observed that two of the three participant gave the first chair of their animation path the highest rate when was asked how willingly they wanted to sit.

The social phobic participants prefered the yellow chair which was the physically the most distance chair and the red chair was twice the fourth preference (see table 3).

Table 3: The preference ranking for the colored chairs for 3 social phobic patients

Participant	Red Chair	Yellow chair	Green chair	Blue chair
1	4.0	2.0	1.0	3.0
2	1.0	3.0	4.0	2.0
3	4.0	1.0	3.0	2.0

CONCLUSION

In order to verify if the prototype virtual environment is recreating the real life situation successfully, and is able to provoke the same human response, two test cases were investigated. The following two test cases are assumed to be representative behavior of social phobia patients, and were used for the experiment with the prototype: (A) People who enter a crowded room want to sit down as quickly as possible to avoid being the center of attention; they would prefer a chair which is close to them, and (B) People have the tendency to physically distance themselves from strangers when choosing a seat in a room full of strangers; they would prefer a chair with no one sitting around them.

The results for research case A showed that the rating of the red chair was higher when given at the beginning of the path, as was the case for animation path 1. Participants following animation path 2, for whom the red chair was

the last one on their path, gave a lower score for the red chair.

The same is observed for the green chair which also gets a higher grade if is graded earlier in the path. For animation path 1, the grades of the green chair are lower than in the case of animation path 2. Therefore we can conclude that the participants prefer to take a seat on the chair that is nearest to them. This is in line with the hypothesed behavior for case A.

Some participants reinforced this conclusion by actually stating that they wanted to sit down as fast as possible, since they did not want to disturb the people already sitting down (avatars). This confirms that they had the feeling of actually being in the virtual room and wanted to sit as soon as possible and not being the center of attention.

In research case B we observed that the red chair was significant more present than the green chair (see Figure 3). Since the only difference with the green chair was the fact that the chairs surrounding the red one were empty, we can conclude that the participants prefer a chair with no avatars around them rather than the chairs with avatars surrounding it.

For our experiment we used students, staff or PhD students of the Delft University of Technology. The Liebowitz Social Anxiety Scale was not taken into account in our results because we only had a limited number (n=3) of participants with a social phobia. Although we observed different results for the social phobia participants, the number of social phobia participants was limited to draw any significant conclusions for that group. The difference in the results for the social phobia participants was that they did not have as strong a preference for the red chair as the other participants. A possible explanation for this behavior could be that the recreated room was not totally symmetric, and although the red chair was not surrounded with avatars, the fact that it was next to a window could have influenced the choice of the social phobia participants. We assumed that the colors used were neutral colors and did not influence the preference of the participant.

Since the assumed behavioral response were observed for both case A and case B, we can conclude that a social situation can be recreated in a virtual world.

REFERENCES

Emmelkamp, P.G.M., M.Krijn, A.M. Hulsbosch, S. de Vries, M.J. Schuemie, en and C.A.P.G van der Mast. „Virtual Reality Treatment versus exposure in vivo;." *A comparative Evaluation in Acrophobia Behaviour Research & Therapy*, 2002.

Hall, E. T. (1968). Proxemics. *Current Anthropology,* 9:83-108.

Heimbert, R.G. and Robert M. Holaway,. 2006. Examination of the known-groups validity of the Liebowitz Social Anxiety Scale." *Depression and Anxiety*, 2006.

Hodges, L.F., P. Anderson, G.C.Burdea, H.G.Hoffman, B.O. Rothbaum. 2001. „Treating psychological and psychological disorders with VR." (IEEE Computer Society Press) 21, nr. 6.

Ickinger, W.J. and Sandra Morris (2001), Psychological Characteristics and Interpesonal Distance, retrieved from http://sharktown.com/proxemics/gfx/media/RSS-PC1.pdf on January 8, 2009.

igroup presence questionnaire (IPQ) Scale Construction. 1995–2004 . http://www.igroup.org/pq/ipq/index.php (retrieved August 21, 2008).

Kessler, R.C. 2003. The impairments caused by social phobia in the general population: -implications for intervention. *Acta Psychiatrica Scandinavica Supplement* 19-27 (2003).

Krijn, M., P. M. G. Emmelkamp, R.P. Olafsson, R. Biemond 2004. Virtual reality exposure therapy of anxiety disorders: A review, *Clinical Psychology Review* 24 (2004) 259-281.

Neal, J.A., Edelmann, R.J. 2003. The etiology of social phobia: toward a developmental profile. *Clinical Psychological Review* 23(6): 761-786. (2003).

Powers, M. and P.M.G. Emmelkamp, 2008. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders, 22, 3*, pp 561-569.

Psychiatric Association, American. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV).* Washington, 1995.

The Anxiety Community. 2000-2005 . http://www.anxietyhelp.org/information/leibowitz.html (retrieved August 20, 2008).

Van der Mast, C. 2006. Technological challenges in the Delft virtual reality exposure therapy system, *Int J Disabil Human Dev 2006; 5(3):205-210.*

Wiersma, J., A. Greeven, E. Berrety, M. Krijn, P. Emmelkamp. „De effectiviteit van Virtual Reality Exposure Therapy voor hoogtevrees in de klinische praktijk." *Gedragstherapie (in dutch)*, 2008: 253-259.

Wittchen, H.U., Fehm, L. „Epidemiology and natural course of social fears and social phobia." (Acta Psychiatrica Scandinavica Supplemen) 108(417): 4-18 (2003).

PRESENCE FOR VR EXPOSURE THERAPY THROUGH 3D ARCHITECTURAL VISUALIZATION

Ervin Sabadi
Willem-Paul Brinkman
Charles A.P.G. van der Mast
Delft University of Technology
Mekelweg 4, 2628 CD Delft
The Netherlands
Email: w.p.brinkman@tudelft.nl

KEYWORDS
Virtual reality, phobia treatment, presence, 3D visualization.

ABSTRACT

This paper reports on the effect of 3D architectural visualization in VR and its possible effect on the sense of presence. The problem definition was tested by use of a between subjects design. Two groups were randomly formed. One group evaluated the 3D visualized virtual environment, and the other the 2D visualized virtual environment. Afterwards the Igroup Presence Questionnaire was used to evaluate the sense of presence. This questionnaire covers three subscales (spatial presence, involvement, and experienced realism) and one additional general item ("sense of being there") not belonging to a subscale.
Two out of three subscales showed a significant improvement of the 3D visualized virtual environment. Involvement and the general item showed no significant difference. It is evident that the 3D architectural visualization provides us with a promising perceptual component for reaching a higher sense of presence for VR exposure therapy.

INTRODUCTION

The motivation for the work presented in this paper should be seen with the context of Virtual Reality Exposure Therapy (VRET). VRET is an instance of behavioral treatment devised on Virtual Reality (VR) technology. VRET has proven its efficacy in treating acrophobia (fear of heights), arachnophobia (fear of spiders), and fear of flying (Krijn et al. 2004). It has also shown promise for the future in treating other phobias like claustrophobia, fear of driving, fear of public speaking, posttraumatic stress disorder (PTSD), and agoraphobia (Powers&Emmelkamp 2008). VR is used to create immersive virtual environments with the aim to expose patients with phobias to controllable levels of anxiety. The patients have to withstand feared situations until their feeling of fear subsides to a certain level before they are directed to more challenging situations. The project "VR and phobias" has been started in 1999 in order to develop a VRET system for treating several phobias (Van der Mast 2006). In an effort to find out more about what causes people to experience the sense of presence (SoP) in VR, a study was started on basis of agoraphobia. Unlike most other phobias, a high avoidance level of feared situations may seriously damage someone's ability to work, to travel, or to even carry out the simplest daily routines. Robillard et al. (2003) indicate a synergistic relationship between presence and anxiety. Brinkman et al. (2008) report a search for parameters evoking presence for social phobia therapy.

Most VR research on the SoP is done by exposing participants to virtual environments. Many of these environments, if not all, are comprised of 2D and/or 3D elements. Even though this is the case, little attention has been paid to differences between the architectual building blocks in the virtual environments as a source for eliciting a SoP. This study focuses on researching 3D visualization in VR and its possible effect on the SoP. Being one of the largest historic market squares of Europe, the Delft market square was chosen for the project "VR and phobias" in researching VRET in respect to agoraphobia.

THEORETICAL BACKGROUND

Agoraphobia

People with agoraphobia have feeling anxiety about being in places or situations from which escape might be difficult (or embarrassing) or in which help may not be available in the event of having an unexpected or situationally predisposed panic attack or panic-like symptoms. Two factors are essential with agoraphobia: *anticipatory anxiety* and *avoidance* of situations that cause anxiety. Anticipatory anxiety is the anxiety experienced by merely thinking about a possible attack, which might occur when starting some activity. It can be severe and even appear hours before the dreaded activity. Avoidance is a behavior which is caused by trying to avoid certain situations or activities, because of the fear of a panic attack. Common themes that accompany agoraphobia are:

- Distance from home
- Traveling alone
- Crowds
- Confinement
- Open spaces
- Social situations

A few example situations to put these themes in a better perspective are as follows:

- Standing in a cue
- Crowded shops
- Empty streets or markets
- Cinemas, theatres

- Traveling by car, train or airplane
- Being in an elevator

Presence

According to literature a defining component of VR systems in general is the feeling "sense of presence" (SoP) (Hodges et al. 1994). In relation to VR the concept of presence is best characterized by transportation; people immersed in VR are thought to feel present in the VE when they have the feeling of 'being there'.
Much research has already been done on the subject of SoP and several researchers (Lombard&Ditton 1997, Witmer&Singer 1998, Slater&Usoh 1993, Sheridan 1992, Steuer 1992) have even composed categorizations of factors that contribute to that feeling.

The tools for measuring the SoP consist of objective measures and subjective measures. Objective measures come in two varieties: behavioral and physiological measures. Behavioral measures are based on behavior a person shows while immersed in a VE (Sheridan 1996, O'Brien at al. 1998). Physiological measures are directed at measuring presence through physiological changes like heart rate, skin temperature and skin conductance (Sheridan 1996). Subjective measures are used most frequently in researching presence. This is done through use of questionnaires. People immersed in a VE are probed with questions related to the projected environment in order to get a better understanding of the concept of presence.

METHODS AND MATERIALS

The idea was to develop two environments that resemble the market square in Delft. One environment would consist of 2D surroundings placed in such a way that they emit the illusion of a 3D setting. The other environment would consist of actual 3D surroundings, built to resemble the real world in a higher degree.
Researching the effect of 3D visualization on the SoP means introducing architecture into the realm of VR. In daily life architecture focuses on design and construction as a means of exhibiting a certain visual experience. Aside from the prominent visual aspect, architecture can also be experienced through our aural, olfactory and tactile senses. As Delft market square was chosen as candidate for this project, this meant recreating the historic market square up to a high level of realism. Using display optics (e.g. a HMD) we create depth perception to our brain by showing our eyes stereoscopic 3D imagery of computer generated visual data.
An experiment with test subjects immersed into the two different virtual environments was held in order to find out if there is a difference in the SoP which can be elicited by the fundamental build-up of the virtual environments.

Requirements analysis

Factors that contribute to a feeling of SoP in respect to 3D visualization in VR are (Lombard&Ditton 1997, Witmer&Singer 1998, Slater&Usoh 1993, Sheridan 1992, Steuer 1992):
- High quality
- Consistency
- Sensory factors
- Distraction factors
- Realism factors
- Vividness

Reference material

In order to start developing the virtual environments of Delft market square, reference materials were necessary. A visit to the market square itself was made. Using a Canon EOS 400D digital SLR camera with standard lens to snap pictures of the market square ended up in an inventory exceeding 200 photographs. The photographs depicted buildings, containers, lampposts, benches, chairs, tables, pillars, and more. All these objects needed to be photographed from different angles as much as possible, because 3D modeling requires it.

Design process

The 3D modeling software package Maya (v7.0.3) from Autodesk was used to model the required objects for the market square in Delft. The photographs formed the basis of each object. Two main categories can be distinguished in the way the objects needed to be modeled. One category consisted of modeling buildings and the remaining category consisted of modeling all other objects. The difference between these two categories lies in the way the original photographs were used. For the first category the
photographs were actually integrated into the final models of the objects, whereas for the second category the photographs were only used as an orientation aid during modeling.

Figure 1: Photograph with perspective

The out-of-camera pictures needed some adjustments before they could be used. The first step in modeling the buildings was to remove the visible perspective (figure 1).

After this the pictures were imported into Maya (and projected onto a polygon plane) where additional adjustments could be made.

The next step in the process was to cut out all the parts of the photograph that were not necessary. This way the

contours of the final 3D model already started to get visible. After that, extra contours needed to be drawn onto the textured polygon plane. This was done around windows, doors, façades, and other components that are subject to visual depth. This can be seen in figure 2.

Figure 2: Building getting contours

Depth was applied onto the surface of the polygon plane on places where contours were drawn earlier in the process (e.g. around windows, doors, façades). This is clearly visible in figure 3.

Figure 3: Depth applied, left: polygon planes, middle: textures added, right: light source added.

The most left representation in figure 3 shows the look of the actual polygon plane without textures applied to it. From the sideway it is clearly visible that depth is present. The middle representation shows the same polygon plane, but now textured with the front view of the building that was being modeled. The depth information is now visible on the textured polygon plane. The most right representation shows the same polygon plane as the middle one, but now with a light source added to the scene.

After the front views of the buildings were modeled, the sides needed to be done as well. The back of the buildings consisted of a mirrored front view. Additional adjustments needed to be made to the roofs, as they needed to be in an angle with the front view. Figure 4 (right side) shows the final result.
After the buildings all the other small objects needed to be modeled as well.

So far the development of the 3D visualized virtual environment. For the 2D virtual environment the same photographs were used for consistency reasons. A 2D building would look like the building depicted in figure 2. The difference with the 3D version of the building would be the lack of depth. Figure 4 (left side) shows a view of the 2D visualized virtual environment.

The view of Delft market square formed the basis for the arrangements of the virtual environments which would be used for the experiment. The market square consisted of a relatively large square surrounded by numerous buildings. As far as the buildings were concerned, an attempt was made to place them in the same consecutive order, resembling the real market square as much as possible. After the Delft market square was modeled in Maya, it needed to be imported in Vizard. Vizard (v3.10.0059) was used to implement interactive navigation in the modeled virtual market square using a Head Mounted Display (HMD).

Figure 4: Final result of the market, left are the 2D models, right the 3D models. For the participants left and right are either 2D or 3D. This picture is only for comparison.

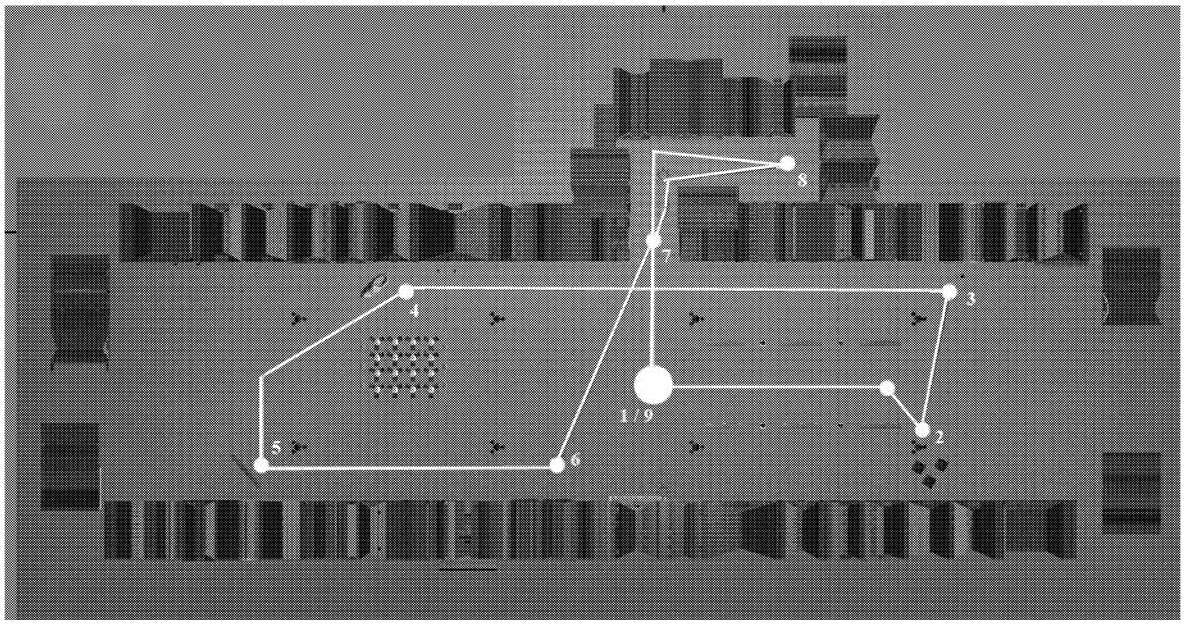


Figure 5: The path to walk.over positions 1-9

THE EXPERIMENT

Experiment set-up

An experiment was designed to test if the hypothesis was supported that the 3D architectural visualization in VR would provide a higher sense of presence in comparison to a 2D version of the same virtual environment. The experiment was run on a system with Microsoft Windows XP (SP3) installed on it. The porting from Maya to Vizard was done through a conversion program PolyTrans (v4.1.2) by Okino. The hardware on which the experiment was run consisted of a Dell Optiplex 755 with a Intel Core 2 Duo E6750 processor, 2048 MB of memory, and a NVIDIA Quadro FX 1700 graphics card. The HMD eMagin Z800 3Dvisor was used for viewing the developed virtual Delft market squares. In the experiment the participants were able to look around freely as the used HMD had a built-in 3 DoF (Degrees

of Freedom) tracker. Because the tracker did not provide sensory information for movements, the navigation was controlled by the leader of the experiment.

The 20 participants were all students from Delft University of Technology. Two groups of 10 participants were randomly chosen. One group evaluated the 3D visualized virtual environment, while the other group evaluated the 2D visualized virtual environment. The average age of the group that tested the 3D VE was 24.1 year, whereas the average age for the group that tested the 2D virtual environment was 23.4 year. All the participants were male. The experiment lasted approximately 8.5 minutes, after which the participants were given the opportunity to assess their virtual environment trial with a questionnaire. The experiment had setup *between-subjects*.

The participants were informed about what was going to happen. They used the HMD to view a virtual environment, after which they would need to assess it by filling in a specific questionnaire to ascertain their SoP. They would be guided automatically between stopping points while they were immersed in the virtual environment, so that they would walk a specific path. The paths that were taken by the participants are depicted in figure 5 by lines following 1,2,3,4,5,6,7,8 and 9.

The numbered dots represent the moments in which the participants were given the opportunity to look around, and are numbered according to the direction taken. There are a few reasons why those points were chosen. First of all when participants entered the virtual environments, they would start at the center of the market square. This way both virtual environments (2D and 3D) got the opportunity to provide maximum depth information (especially the 2D virtual environment provides maximum depth information when a participant is right-angled towards the buildings). A few stops were made around objects, as extra visual depth information should be available at those places. This would provide both modeled virtual environments with the same kind of chance of acquiring a feeling of being there. Two trajectories were next to the buildings, in order for the participants to pick up the details brought on to the various buildings available in the 3D virtual environment, but missing in the 2D virtual environment. Hopefully this would trigger a difference between the two virtual environments in respect to a SoP. At point six in figure 5 a diagonal path was chosen directed towards the alley available in both virtual environments. The diagonal approach should favor both virtual environments with some points on depth information (even though the 2D virtual environment did not have 3D buildings, the HMD would provide a stereoscopic (depth) view of the alley), although it was expected that the 3D would reach higher sense of presence. Right at the beginning of the alley a stop was organized. The participants were given a chance to look around here and look at the buildings from close by. The 3D virtual environment should have an extra element of realism here, because of the 3D aspects of the buildings from nearby. After this the path continued into the alley, where it was the first time the participants were totally enclosed by buildings. The last trajectory was to walk out of the alley towards the starting point. This trajectory was chosen because of the changing view while leaving a narrow alley. It should give a good sense of presence in the virtual environments.

A total of eight paths had been introduced for the participants to take, and nine positions in which they would be left stand still in order for them to get a feel of their surroundings. At the stopping positions (including the starting position) the participants would get 30 seconds to look around freely before moving on to the next one. In total they would spend 235 seconds 'walking' (in which they would also be able to look around), and 270 seconds 'looking' (while stationary). The timings of each path the participants were taken, are documented in table 1. As a total, the participants were immersed into a virtual environment for approximately 8.5 minutes.

Table 1: Timing of the paths.

1.	*Path 1 → 2*: 30 s
2.	*Path 2 → 3*: 15 s
3.	*Path 3 → 4*: 50 s
4.	*Path 4 → 5*: 30 s
5.	*Path 5 → 6*: 25 s
6.	*Path 6 → 7*: 30 s
7.	*Path 7 → 8*: 20 s
8.	*Path 8 → 9*: 35 s

The questionnaire used to evaluate the SoP of the participants in the two different virtual environments, is the Igroup Presence Questionnaire (IPQ). The IPQ consists of 14 items, covering three subscales (spatial presence, involvement, and experienced realism) and one additional general item ("sense of being there") not belonging to a subscale (IPQ, 2008).

RESULTS

The IPQ is based on a 7-point Likert scale itemts. This means that the answers can range from -3 (e.g. fully disagree) to +3 (e.g. fully agree), whereas position 0 portraits the meaning of neutrality towards the given situation. In order to make certain that the three subscales were consistent for this type of experiment, the Cronbach's alpha was determined. For spatial presence $\alpha = 0.83$, for involvement $\alpha = 0.64$, and for experienced realism $\alpha = 0.80$.
To present the data in a way that distortion factors are excluded as much as possible the Likert scales were reduced to an ordinal level. All the responses were combined into three categories (negative/ neutral/positive). This means that the Likert scales

were relabeled as follows: [-3, -1] as negative, ⟨-1, 1⟩ as neutral, and [1, 3] as positive towards a SoP.
Examing the graph two IPQ subscales (*spatial presence* and *experienced realism*) showed a clear difference for VR (2D or 3D) in eliciting a SoP. Concerning *spatial presence* it is visible that the large majority of the participants stated that the 2D visualized virtual environments had a neutral effect on the SoP, whereas the participants who were immersed in the 3D visualized VE stated more often a positive effect on the same matter. The same applies for *experienced realism*, where the 3D visualized virtual environments shows better results as well. The reactions from the participants who were immersed in the 2D visualized virtual environments can mostly be found in the negative/neutral area, whereas the 3D visualized VE shows overwhelming responses in the neutral area. The results are presented in Figues 6 and 7.

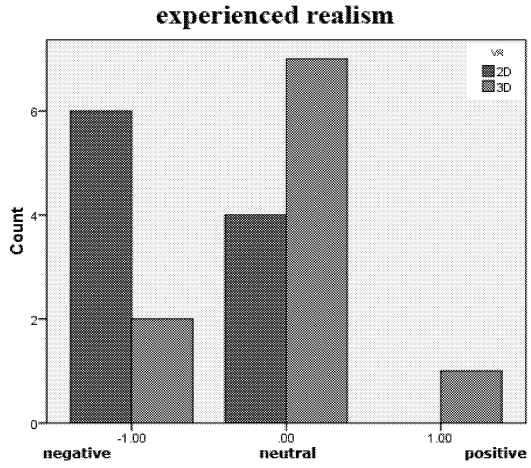


Figure 6: Experienced realism

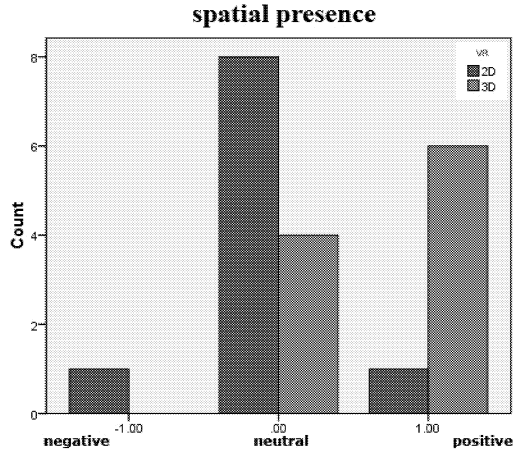


Figure 7: Spatial presence

Mann-Whitney U tests confirm these differences as can be seen in Table 2.

Table 2: Results of Mann-Whitney U

Measure	*p*	*U*	*Z*
Spatial presence	0.018	23.0	-2.37
Involvement	0.264	37.5	-1.12
Experienced realism	0.058	28.0	-1,90
'Being there'	0.166	34.5	-1.39

Subscale *experienced realism* reached a value close to an α level of 0.05, that we could practically tag it as significant. Moreover, with a p = 0.018, we were able to add subscale *spatial presence* on the list of significance. This means that out of three available subscales we were able to find two of them showing significant differences between the 2D and 3D visualized VEs.

CONCLUSIONS AND RECOMMENDATIONS

Two out of three subscales showed a significant improvement of the 3D visualized virtual environment. One subscale and the general item showed no significant difference. Having found significant differences between the 2D and 3D visualized virtual environment is quite remarkable considering the small test group used in the experiment. It is evident that the 3D architectural visualization provides us with a promising perceptual component for reaching a higher sense of presence. It would be interesting to know what would happen with a larger test group.

A way to emphasize architecture in a three dimensional form in virtual reality could be done by using active lighting. Inserting active lighting in virtual reality would make it possible to introduce shadows. For virtual reality systems optics are used to fool the brain by showing our eyes depth perception. In our experiment architecture was used to actually introduce real depth in virtual reality, which showed promising results. By emphasizing the architectural factor in virtual reality through use of shadows, we may be able to further enhance the sense of presence. Vizard, the interactive real-time 3D program which was used for our experiment, does not support active shadowing through the introduction of lights. But perhaps it will be available in a future update of the program.

The people who participated in the experiment used in this research were all male (19 to 34 years) and technically educated. It is conceivable that a different composition of the participants and different age groups may result in different findings. We can expect that in the future patients who suffer phobia have a lot of experience with computer games and other high quality 3D worlds. And probably more 3D-realism might be required to evoke presence for them.

REFERENCES

Brinkman, W.-P., C.A.P.G. van der Mast, and D. de Vliegher. 2008. Virtual reality exposure therapy for social phobia: A pilot study in evoking fear in a virtual world, *Proceedings of HCI2008 workshop - HCI for technology enhanced learning*, ISBN 978-90-813811-2-3, 85-88, Delft University of Technology, Mediamatica, Delft, The Netherlands.

Hodges L., B.O. Rothbaum, R. Kooper, D. Opdyke, T. Meyer, J.J. De Graaf, J.S. Williford. 1994. Presence as the defining factor in a VR application, *Technical report GIT-GVU-94-5*, Georgia Institute of Technology.

IPQ. 2008. Igroup Presence Questionnaire Scale Construction. http://www.igroup.org/pq/ipq/index.php. (retrieved August 21, 2008).

Krijn, M., P. M. G. Emmelkamp, R.P. Olafsson, R. Biemond. 2004. Virtual reality exposure therapy of anxiety disorders: A review, *Clinical Psychology Review* 24 (2004) 259-281.

Lombard M, T. Ditton. 1997. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2). Available: http://jcmc.indiana.edu/vol3/issue2/lombard.html

O'Brien J, M. Büscher, T. Rodden, J. Trevor. 1998. Red is behind you: The experience of presence in shared virtual environments. Paper presented at the *Workshop on Presence in Shared Virtual Environments*.

Powers, M. and Emmelkamp, P.M.G. 2008. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders, 22, 3*, 561-569.

Robillard, G., S. Bouchard, T. Fournier and P. Renaud. 2003. Anxiety and Presence during VR Imeersion: A Comparative Study of the Reactions of Phobic and Non-phobic Participants in Therapeutic Virtual Environments Derived from Computer Games, CyberPsychology & Behavior, Vol. 6, No. 5, 467-475.

Slater, M., M. Usoh. 1993. Representation systems, perceptual positions, and presence in immersive virtual environments. *Presence*, 2: 221-233.

Sheridan T.B.. 1992. Musings on telepresence and virtual presence. *Presence*, 1: pp 120-126.

Sheridan T.B.. 1996. Further musings on the psychophysics of presence. *Presence*, 5: 241-246.

Steuer J. 1992. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4): 72-93.

Van der Mast, C. 2006. Technological challenges in the Delft virtual reality exposure therapy system, *Int J Disabil Human Dev 2006; 5(3):205-210.*

Witmer, B.G., M.J. Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7: 225-240.

DATA TRANSMISSION AND SECURITY

A High Performance Interference Canceller with Narrow Input for Carrier Superposed Satellite Communication

Takehiro Ishiguro, Shoko Kuroda, Sho Tanaka,
Ryusuke Miyamoto, Takao Hara, and Minoru Okada

Dept. of Information Systems,
Graduate School of Information Science,
Nara Institute of Science and Technology,
8916-5, Takayama-cho, Ikoma, Nara, Japan

KEYWORDS
Superposed Transmission, Satellite Communication, Limited Data Width

ABSTRACT

A carrier superposed transmission technique is a promising method for improving frequency utilization e ciency of a satellite communication system. In order to realize the carrier superposed transmission, exetended matched lter (EMF) based propagation delay measurement method has been proposed by the authors. However, width of input signals to canceller via A/D converter is not discussed in the EMF-based scheme, which a ects the performance of cancellation. In this paper, we proposes a novel cancellation scheme which adopts demodulation and FIR lter compensation before EMF matching. Experimental results shows that the proposed scheme achieves highly accurate cancellation under narrow input data width caused from system constrains and with IB signal of any symbol rate.

Introduction

Superposed transmission technique is a technique to improve frequency utilization e ciency of bi-directional satellite communication systems. In this technique, both out-bound (OB) and in-bound (IB) signals are superposed by assigning them to the same frequency band contrary to the conventional systems which assign those signals to di erent frequency bands.

In satellite communication, two types of bi-directional systems are available; Point to Multi-Point (P-MP) system and Point to Point (P-P) system. The con gurations of P-MP and P-P systems are described as Fig.1 (a) and (b) respectively. P-MP system is composed of a central hub station having a large aperture antenna and many very small aperture terminals (VSAT). From this, P-MP system is also called as VSAT system. On the other hand, P-P system has two stations on earth having a similar sized antenna and are communicating via the satellite. In both systems, receiving signals are the sum of signals from all stations on earth since all signals are transmitted at the same frequency band. In such case, desired signals are interfered by the signals transmitted from the station itself. Therefore, to demodulate the desired signal correctly, an interference must be eliminated by cancelling unnecessary signals from the received signal.

The authors has developed the interference canceller suitable for the superposed transmission on P-MP system(Hara et al., 2005; Kobashi et al., 2006; Osato et al., 2007a,b). However the canceller for P-MP system is not applicable to P-P system because unlike the P-MP systems, the signal power strength of both stations on earth are about the same in P-P system.

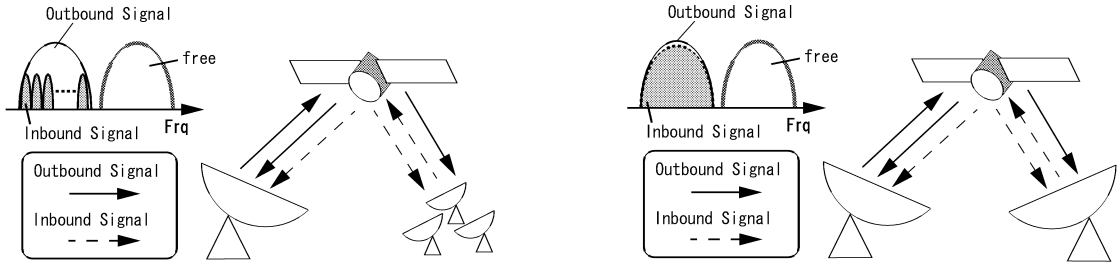
To realize the canceller suitable for a P-P system, the authors proposed a cancellation scheme based on Extended Matched Filter-based (EMF)(Tanaka et al., 2008). However, width of input signals to canceller via A/D converter is not discussed in the EMF-based scheme, which a ects the performance of cancellation. If data width of input signals are limited by system constrains, su cient performance cannot be achieved. To solve such problem, we propose a novel canceller that achieves su cient performance even if data width of input signals is small.

The rest of the paper is organized as follows. The next section details the EMF-based canceller proposed in (Tanaka et al., 2008). Third section proposes a high performance interference canceller with narrow input for carrier superposed satellite communication. Experimental results in order to con rm the validity of the proposed method are shown in fourth section. Finally the last section concludes this paper.

Existing Interference Canceller for Satellite Communication

The concept of a canceller proposed in (Tanaka et al., 2008) is shown in Figure 2. In Figure 2, two stations transmit the signals, the signals to the left and right are respectively denoted by $s_1(t)$, and $s_2(t)$.

Let us assume that the cancellation is performed at the



(a) a P-MP system (b) a P-P system

Figure 1: Carrier allocation of P-MP and P-P bi-directional satellite communication systems with superposed transmission

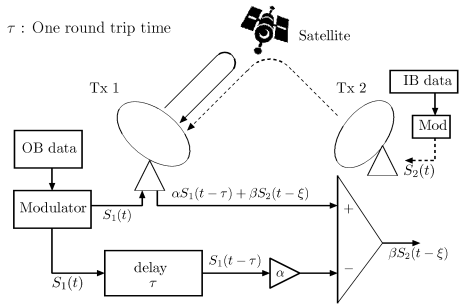


Figure 2: Concept of interference canceller for superposed transmission

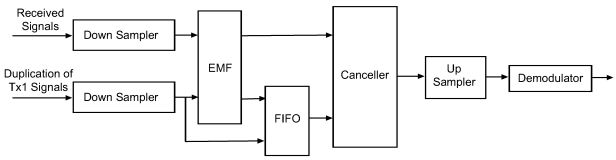


Figure 3: Block diagram of an existing interference canceller with EMF

rst earth station or the station to the left in the Figure. Since both the earth stations transmit the signal at the same frequency, the received signal is the sum of $s_1(t)$, and $s_2(t)$. The received signal at the rst station is given by:

$$r(t) = \alpha s_1(t-\tau) + \beta s_2(t-\xi) + z(t), \qquad (1)$$

where α and τ are the round-trip propagation path loss and round trip delay between the rst earth station and the satellite, and β and ξ are the propagation path loss and delay between earth stations via satellite. Furthermore, $z(t)$ is an additive white Gaussian noise(AWGN) component.

At the canceller, the transmitted signal, $s_1(t)$, is applied to the delay block whose delay time is set to be the same as the round trip delay, τ. The output of the delay block $S_1(t-\tau)$ is then fed to the variable gain ampli er to adjust amplitude of the replica and the interference signal. If the adjustment is perfect, the gain of the variable gain ampli er is set to be α. The output of the variable gain ampli er is then subtracted from the received signal. The output of the interference canceller is given by

$$u(t) = r(t) - \alpha s_1(t-\tau) = \beta s_2(t-\xi) + z(t). \qquad (2)$$

Figure 3 shows a detailed block diagram of the EMF-based interference canceller proposed in (Tanaka et al., 2008). The received signal from the satellite and the transmitted signal from the transmitter is converted to equivalent low pass expressions and applied to the canceller. The sampling frequency of these signals are adjusted at the rate converters (RateConv) as twice as the symbol rate of the transmitted signal. The two signals are then applied to EMF. The transmitted signal is also stored to FIFO (First-In First-Out) memory to delay the signal.

Proposed Interference Canceller with Narrow Input for Carrier Super-posed Satellite Communication

The existing EMF-based scheme can minimize error between the replica signal and the received Out Bound (OB) signals by controlling phase and amplitude properly. However, accuracy of feedback error is degraded if data width of input signals become small. To improve performance of cancellation when data width of input signals is not wide enough, we propose a novel canceller that has demodulator and modulator before EMF in Figure 3.

In the proposed scheme, input signals are once demodulated after down sampling and then the signals are modulated by the same method used for the modulation of input signals. In these processes, degradation of the input signals caused by insu cient width of the data path

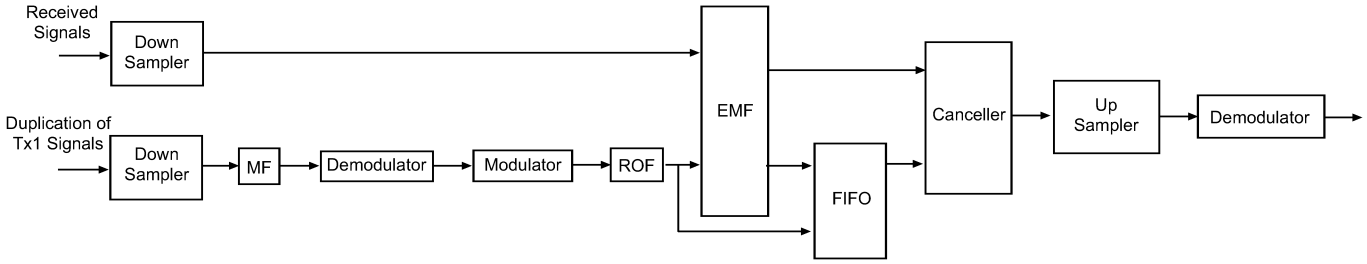


Figure 4: The Proposed Interference Canceller

is reduced. After the re-modulation, roll-o lter is applied. After above processes, the signals are prosecced as same as the existing scheme shown in Figure 3.

Evaluation

In this section, we evaluate the performance of the proposed scheme by computer simulation.

Simulation Parameters

In this evaluation, QPSK modulated signals at the same band are adopted for both IB and OB signals. The power of these signals are the same and maximum valid size of data width is 10 bits. Valid bit width of Roll O Filter (ROF) is 9 bits.
The other system parameters were set as follows. The number of EMF taps is 64 and the clock frequency is 72 MHz. The symbol rate is 10 Msps and the over sampling factor is 2.

Performance of cancellation

To evaluate the performance of canceller when input data width is narrow or su ciently wide, we varied the input data width from 4 bits to 10 bits. The results of cancellation with no IB signal are shown in Figure 5 and Figure 6. As shown in these gures, it cancels unnecessary signals better in no IB signal condition without the demodulation before the EMF when the input data width is 10 bit but it cannot cancel properly without the demodulation when the data width is 4 bit. The results of cancellation with 2.5 Msps IB signal are shown in Figure 7 and Figure 8. As shown in these gures, 4 bit input with no demodulation has degraded performance compared to the ones with demodulation. For 10 bit input, both conditions have about the same performance. The results of cancellation with 5 Msps IB signal are shown in Figure **??** and Figure **??**. As shown in these gures, 4 bit input without demodulation cannot cancel unnecessary signals properly and with 10 bit input, both conditions have about the same performance.
According to these results, our proposed method has capability of cancelling unnecesary signals under narrow input data width caused from system constrains and with IB signal of any symbol rate.

Conclusion

In this paper, we proposed a novel cancellation scheme that adopts demodulation and roll-o lter compensation before EMF matching. Experimental results showed that the proposed scheme achieves accurate cancellation under narrow input data width caused from system constrains and with IB signal of any symbol rate.

REFERENCES

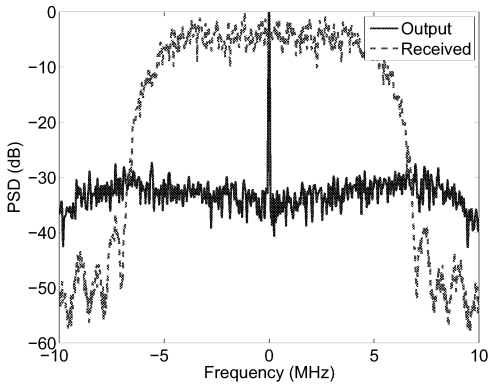
Hara T.; Ichikawa M.; Okada M.; and Yamamoto H., 2005. *Canceler Design for Carrier Super-Positioning Frequency Re-use of VSAT Satellite Communications. IEICE Transactions of Communications*, J88-B, no. 7, 1300–1309.

Kobashi H.; Osato M.; Hara T.; Okada M.; and Yamamoto H., 2006. *Behavior of the symbol error of replica in the canceler for VSAT frequency reuse satellite communications.* In *Proc. of IASTED Conference on Wireless Networks and Emerging Technologies (WNET 2006).* 311–316.

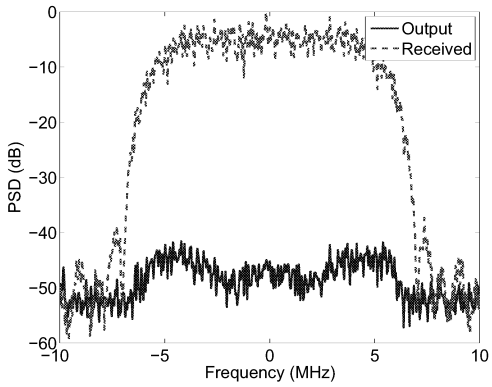
Osato M.; Hara T.; and Okada M., 2007a. *Stable Delay Measurement System for Satellite Carrier Super-Positioning. IEICE Transactions on Communications*, J90-B, no. 12, 1301–1313.

Osato M.; Kobashi H.; Omaki R.Y.; Hara T.; and Okada M., 2007b. *Signal Canceler in the Carrier Superpositioning Satellite networks. Journal of Communications*, 2, no. 7, 68–76.

Tanaka S.; Kuroda S.; Miyamoto R.; Hara T.; and Okada M., 2008. *A matched filter based round-trip delay measurement method for bi-direc tional satellite communication system with superposed transmission.* In *Proc. of Joint Conference on Satellite Communications (JC-SAT2008).*

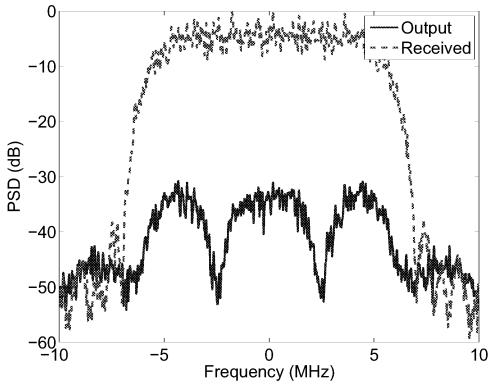


(a) Width of 4 bits

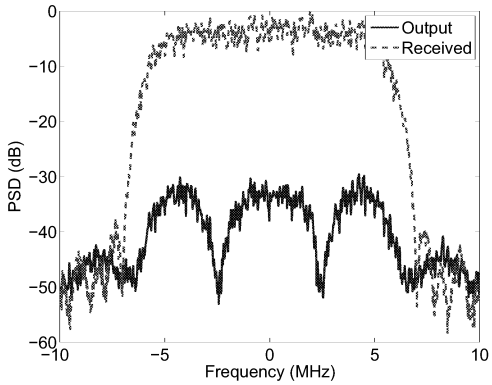


(b) Width of 10 bits

Figure 5: Results of cancellation without demodulation receiving no IB signals.

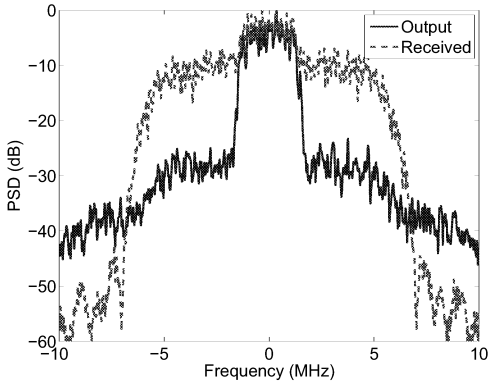


(a) Width of 4 bits

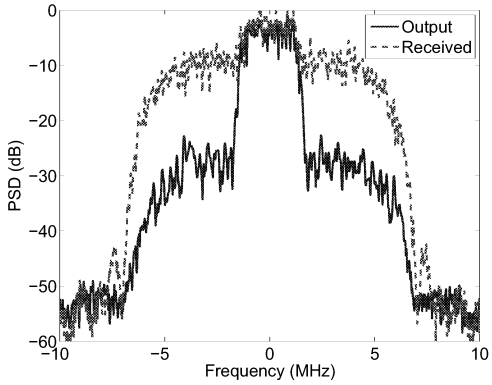


(b) Width of 10 bits

Figure 6: Results of cancellation with demodulation receiving no IB signals.

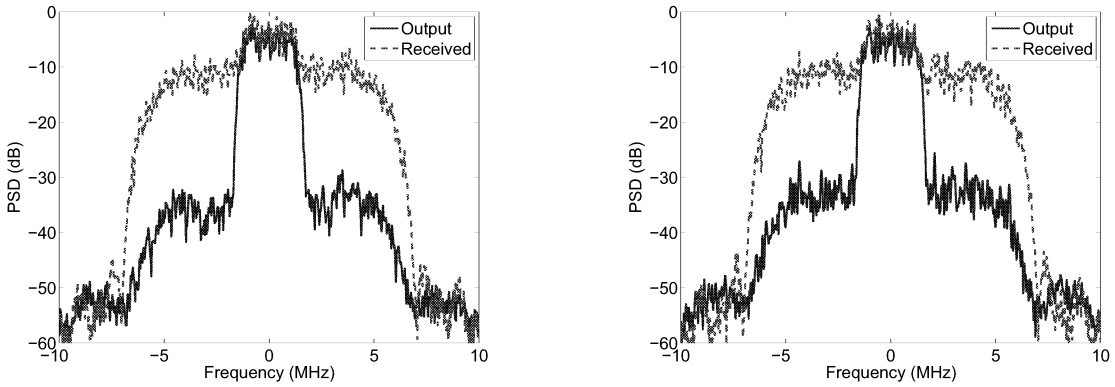


(a) Width of 4 bits



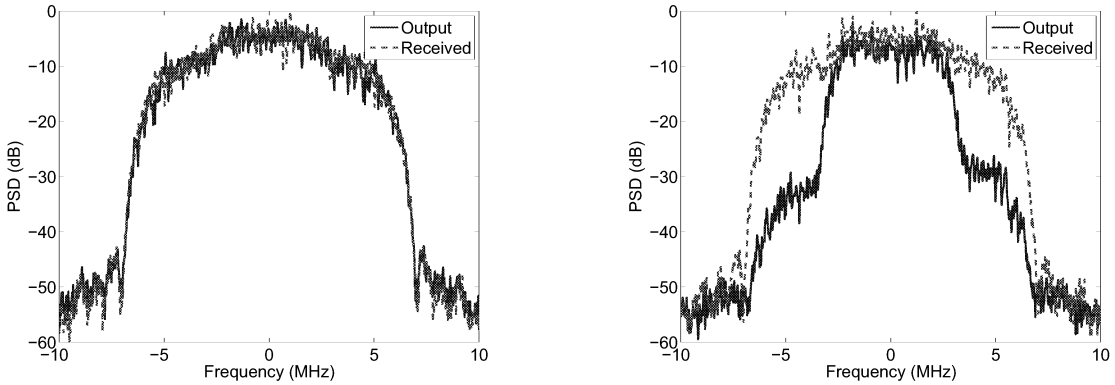
(b) Width of 10 bits

Figure 7: Results of cancellation without demodulation receibing 2.5 Msps IB signals.



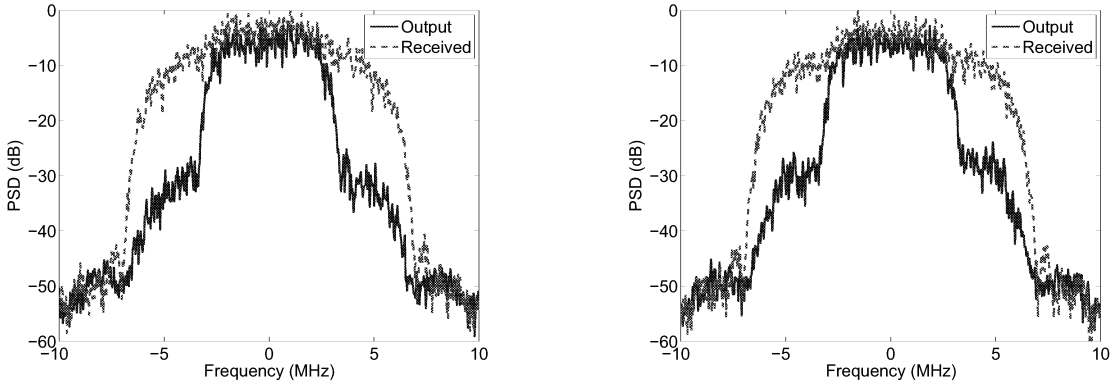
(a) Width of 4 bits (b) Width of 10 bits

Figure 8: Results of cancellation with demodulation receiving 2.5 Msps IB signals.



(a) Width of 4 bits (b) Width of 10 bits

Figure 9: Results of cancellation without demodulation receiving 5 Msps IB signals.



(a) Width of 4 bits (b) Width of 10 bits

Figure 10: Results of cancellation with demodulation receiving 5 Msps IB signals.

COMPARISON OF ATTACKS ON IPV4 AND IPV6 PROTOCOLS

Mojca Ciglarič, Andrej Krevl, Matjaž Pančur
University of Ljubljana, Faculty of Computer and Information Science
Tržaška 25, 1000 Ljubljana,
Slovenia
E-mail: Mojca.Ciglaric@fri.uni-lj.si

KEYWORDS

Network security, IPv6, attack, spoofing

ABSTRACT

The paper overviews the main security-related network attacks and gives the attackers' motivation. Since the paper focuses on the differences between IPv4 and IPv6, the attacks are grouped accordingly. The paper gives an overview of sniffing, application layer attacks, rogue devices, man-in-the-middle attacks and flooding; inquiring, unauthorized access, fragmentation and header manipulation, IP spoofing, ARP and DHCP attacks, broadcast amplification attacks (smurf), routing attacks, viruses and worms. For each of them, a general execution idea and implementation are discussed in both protocols and recommendations for detection or prevention are given.

INTRODUCTION

In IPv6 protocol, many security- related problems are the same as in IPv4 protocol since the basic mechanisms for datagram transfer remain unchanged. However it is a common opinion that IPv6 is a more secure protocol than IPv4, since at least the use of IPSec mechanism is obligatory in IPv6. This assumption is only true to the certain extent: in an idealized environment, with well-coded applications and a robust infrastructure. But in a real life, networks with both protocols are susceptible to similar types of attacks. Most of the security issues are found at the application layer and therefore IPSec cannot guarantee enough security.

In IPv4 networks, firewalls and gateway routers are responsible for several security-related functions. They can also coexist within one device. Firewall policy usually allows outbound traffic while limiting inbound traffic. Additional security elements may be used – intrusion detection and prevention systems as well as application proxies. Usually, publicly accessible servers are placed into an isolated network segment, called Demilitarized Zone. Similar ideas are implemented in IPv6 networks.

The remaining of this paper consists of two parts: the first part describes some well-known network attacks which are feasible in IPv4 as well as in IPv6 in an very similar fashion. The second part describes the attack types in IPv6 that are different from those in IPv4. Although we cannot describe all of them in detail, we give a general idea and pointers to further sources of information. Throughout the paper, we give recommendations on prevention or at least detection of such attacks and therefore we believe the paper will be of interest to a wider audience; to network administrators, to end users and also to everyone else contemplating the IPv6 deployment in their network.

ATTACKS WITH STRONG SIMILARITY BETWEEN IPv4 AND IPv6

Some attacks have almost identical anatomy either we study them in IPv4 or in IPv6 based network. These belong to five types: sniffing, application layer attacks, rogue devices, man-in-the-middle attacks and flooding. They are either not directly related to the network layer protocol or based on the inherent network layer properties - independent of the protocol. Let us briefly review each of them.

Sniffing

The word sniffing denotes a series of attacks, where the attacker intercepts the data while traversing the network. The tools are called network sniffers and are widely available. A very common sniffer is, for example, Tcpdump, a command line tool included in most Unix-based operation systems. The attacker using a sniffer may intercept all kinds of data transferred via unencrypted text-based protocols, such as POP3 or SMTP. Use of IPSec is a good protection against such attacks (the intercepted data is not readable), however its key management protocol is not exactly easy to use. The attack is performed on the network layer.

Application Layer Attacks

Most of today's internet attacks fall into the category of application layer attacks. The application vulnerabilities

account for the majority of modern networks threat sources. IPv4 as well as IPv6 protocols are not really related to this kind of attacks, although with the use of strict IP address authentication some of the attacks would be easier to track back. However, the main causes are found at the application layer. Even an encrypted connection cannot prevent from application layer attacks, neither can limit the damage. In case of attack, the main benefit of encryption is easier tracking back to the source of attack due to the authentication - when we can be sure that the network layer information couldn't have been spoofed.

Rogue Devices

All of the devices which are not authenticated can be viewed as rogue devices, for example a plain old notebook. However the attackers are primarily interested in wireless access points, DHCP servers, DNS servers, routers or switches. When an attacker establishes his own device which then pretends to be a regular network device, we call it a rogue device. For example, when establishing rogue DHCP server and flooding the regular DHCP server, the attacker can lease IP addresses from his own address pool to the network users and then route their traffic through his own infrastructure, where he can freely access it and thus perform man-in-the-middle attack.

Man-in-the-middle Attacks

Neither IPv4 nor IPv6 headers enable any security mechanism. Both of them rely on IPSec protocol, while IPSec is equally prone to the attacks in both cases, with IPv4 and IPv6. The weakest point is in its IKE (Internet Key Exchange) protocol, which enables setting up a security association: with Diffie-Hellman key exchange, a shared session secret is established and used to extract cryptographic keys. IKE has known vulnerabilities to man-in-the-middle attacks, which are propagated in IPSec.

Flooding or DOS Attacks

With larger address space, the identification of DOS attacks in IPv6 is not any easier, but the basic idea of attack is the same as in IPv4. DOS and DDOS network attacks are all about flooding a network device or network connection with more traffic than it can possibly handle. This is the easiest and a guaranteed way to succeed in interruption of its normal operation.

ATTACKS THAT ARE DIFFERENT IN IPv6

Now, let's take a look at the interesting part. With IPv6 deployment, several types of attacks are changed, some security holes are sealed and a few new ones emerged. Some attacks are more complicated in IPv6, the others are easier and in the last group, just the means of attack has changed. We will take a look at the following: inquiring, unauthorized access, fragmentation and header manipulation, network and other layer spoofing, ARP and DHCP attacks, broadcast amplification attacks (smurf), routing attacks, viruses and worms.

Inquiring and Sniffing

The attacker usually tries to get as much information as possible about the target network that he wants to attack. Passive methods include collecting information via search engines and public c documents, while an active method is for example port scanning. In IPv4 networks, the tools and methods are the following:

- Ping: the attacker tries to identify the active nodes on the target network and build a logical network map. Beside ping, other tools may be used (traceroute, firewalk…)
- Port scanning: when an attacker identifies active systems, he may systematically scan the ports to find active and reachable services and use them.
- Application vulnerabilities search: the attacker determines the host operating system version and applications' version numbers.

Several tools are available for easier execution of these actions: Nmap (network mapper), Amap, Nessus…

In IPv6, inquiry techniques are similar to IPv4 with two differences: IPv6 hinders observations on the number of active hosts in a subnet, while new multicast IPv6 addresses enable the attacker to locate systems of certain types (routers, servers) easier.

IPv6 default subnet address size is 64 bits. If the attacker can scan a million of addresses in a second, he may scan an IPv4 subnet in a few seconds, while he would need roughly 500.000 years to scan all of the available IPv6 subnet addresses. In a typical network with a hundred hosts, statistics says that he would find the first active address after 2800 years. However, there are a few facts that make scanning easier: DNS is used for all internet-accessible services, often even for all the network hosts. On the other hand, administrators tend to assign easy to remember IPv6 addresses to the key systems, for example ::10 or ::20. The attacker may also make use of an unprotected router to peek in a neighbour cache and find out accessible host addresses.

IPv6 multicast addresses are defined as follows: FF05::2 is a site-local multicast router address and FF05::3 is a site-local DHCP server address. Knowing that, an attacker

might obtain all the router addresses or all of the DHCP server address and attack with flooding. An important countermeasure is disabling access to these addresses from the outside network.

In IPv6, the attacker can make use of similar tools as in IPv4 network: Nmap (for linux-based systems), alive6-local, alive6-remote, halfscan6 etc.

To prevent IPv6 scanning, the following measures are advisable.

1. Border routers should filter traffic to internal (site-local) IPv6 addresses (i.e. to multicast router and DHCP server addresses).
2. Administrators should assign less obvious static IPv6 addresses to the critical systems.
3. Unused IPv6 service addresses should be filtered out at the firewall.
4. Selective ICMP filtering: Neighbor discovery protocol uses ICMP v6 messages (NA, NS, RA, RS), which should be allowed, while most other ICMP messages may be filtered out (with exception of Packet-too-big, which is used for determining the MTU size). The example belov shows how to filter out unneeded ICMP traffic:

```
permit icmp any-any nd-na
permit icmp any-any nd-ns
permit icmp any-any packet-too-big
deny icmp any-any
```

5. Secured hosts and applications: applying the security patches and maintaining the security level.

Unauthorized Access

By design, in an IP network it is allowed to establish a connection between any pair of network nodes. When an attacker takes advantage of this fact, we talk about unauthorized access. In both IPv4 and IPv6, the administrators need to take care of unauthorized access by using access control techniques at the network as well as at the transport layer. At the network layer, access control lists permit or deny access to resources and they are implemented either on the network devices (firewall) or at the end points (i.e. firewall within the host operating system). Network layer policy filters traffic on the basis of the IP header contents, while more detailed control is possible when making use of deep packet inspection – observation of the higher layers' control data within datagram body.

In IPv6, access control is needed from the same reasons, although a part of it is already handled by IPSec protocol. The first difference is in addressing: in IPv6, a single interface may have several IPv6 addresses (link-local, site-local and global unicast), however global unicast addresses should only be assigned to the devices that need to be publicly accessible. In IPv6 networks, access control is still implemented by firewalls and access control lists, but administrators should take into account several factors when designing security measures:

1. When **IPSec** is in use, cryptographic protection only allows the firewall to inspect data at the network layer. However although IPv6 only makes use of the authentication header, firewall can still control upper layer headers.
2. **Header extensions**: in IPv6, one of the optional header extensions is routing header. Since IPv6 hosts should not drop a datagram containing routing header extension, an attacker could use this property to sneak in his traffic. The organization should make sure that inside traffic does not contain routing headers. Alternatively, when this is not applicable, access control devices should have filtering enabled.
3. **ICMP** filtering should be as strict as possible. Table 1 shows the list of ICMP v4 and ICMPv6 messages that should be allowed to pass through the firewall, Fragmentation related messages in IPv4 are substituted with Packet too big in IPv6, while Parameter problem indicates difficulties with header field identification in IPv6 header or header extension.

ICMPv4	**Meaning**	**ICMPv6**	**Meaning**
Type 0	Echo reply	Type 129	Echo reply
Type 3 code 0	Destination unreachable	Type1 code 0	no route to destination
Type 3 code 4	Fragmentation Needed and Don't Fragment was Set	-	
Type 8	Echo request	Type 128	Echo request
Type 11	Time Exceeded	Type 3	Time Exceeded
		Type 2	Packet too big
		Type 4	Parameter problem

Table 1: ICMP Messages to be Allowed Through the Firewall

ICMPv6	Meaning	Comment
Type 2	Packet too big	Needed when determining MTU
Type 130-132	Multicast listeners	IPv6 routing device should accept these messages in order to take part in multicast routing
Type 133-134	Router Solicitation and Router Advertisment	Important for IPv6 self-configuration, but also for other purposes
Type 135-136	Neighbor Solicitation and Neighbor Advertisment	Useful for duplicate address detection and MAC address resolution.
Type 4	Parameter problem	

Table 2: ICMPv6 Messages to be Allowed in and out of the Firewall

Table 2 shows further IPv6 messages that should be allowed in and out of the firewall. All other ICMP messages may be filtered out.

6. **Multicast control.** Most IPv4 firewalls employ only minimal multicast control. In IPv6, local multicast is integral part of operation and firewalls should not block local multicast addresses to facilitate neighbor discovery protocol. Firewalls should not forward link layer multicast addresses, they should control source IPv6 addresses and filter the datagrams with multicast source address.
7. **Firewalls**. In IPv4, firewall has to recognize IP and link layer interactions such as ARP and DHCPv4. IN IPv6, firewalls have to control corresponding ICMPv6 and multicast messages to facilitate neighbor discovery, duplicate address detection, autoconfiguration, multicast management etc. Multicast-related security policy is to be defined separately. Proper firewall configuration can also diminish consequences of attacks on MAC address - IP address pair and DHCP-related attacks.

To protect the network from unauthorized access, we recommend the following:

- Determine acceptable IPv6 header extensions. For example, the datagrams containing routing header are filtered out with the following line:

```
deny ipv6 any any routing
```

- Determine the necessary ICMPv4 and ICMPv6 messages (see Tables 1 and 2).

Fragmentation and Header Manipulation

In IPv4, fragmentation is a means to adapt large datagrams to smaller MTU on the way to the destination. But it is also a means to avoid security mechanisms, for example IDS, IPS and firewall inspection. Together with other header manipulations, it can also be used for a direct attack on network infrastructure. As a first sign: a high percent of fragmented traffic in IPv4 is an indication of attack or at least attempt. Attackers can make use of tools such as *whisker* or *fragroute*. A common attack is teardrop – sending invalid IPv4 fragments with overlapping, over-sized payloads to the target machine. Stateful firewalls can detect and drop such fragments. If overlapping fragments reach the target operating system, IP stack might still detect and drop them. If not, there is no other barrier. Attacker can also deceive IPS by sending out of order fragments.

In IPv6, fragmentation is only allowed at the end nodes. Also the lowest MTU size is defined at 1280 bytes. If the sending operating system creates fragments in the size determined from the PMTUD messages, network devices may drop shorter fragments, unless they represent the last fragment. If the sending operating systems behave differently, the network may experience a lot of fragmented datagrams and consequently higher risk for fragmentation related attacks.

The following are some recommendations for reducing the effects of fragmentation related attacks:

- Drop the fragments with the destination address set to addresses of the network devices.
- Sufficient capabilities for advanced filtering of fragmented datagrams: security devices should join the fragments to enable detailed analysis of the contents.
- Drop the fragments, smaller than 1280 bytes, with exception of the last one.

IP Spoofing

Several types of network attacks are based on forged or spoofed IP addresses. The recipient thinks that the traffic is sent from somewhere else. In IPv4, such attacks happen on a daily basis and hold back the efforts of tracking to the source of worms, viruses, spam and DoS. Making use of network layer filtering, we might for example detect the datagrams with spoofed local network addresses, coming from the outside, however we cannot detect spoofed local

addresses originating from inside or having only the host portion of the address spoofed.

In IPv6, the address assignment may be checked by several actors within the network. For example, internet service providers could make sure that their clients do not spoof IP addresses; however this is not a standard or required activity to be expected from an ISP. IP address spoofing can be prevented to the certain extent by using access control lists.

Spoofing at the transport layer is the same in IPv4 as in IPv6 since transport protocols are the same. Spoofed traffic may be detected by firewalls (IPv6 enabled) or IDS.

Recommendation for transport layer spoofing and IP spoofing detection and prevention:

- Traffic filtering (performed by users as well as ISPs).
- Documented backtracking procedures. With so many IPv6 addresses that may be spoofed it is very important that we are able to determine the real source of network traffic.
- Use of cryptographic protection.

ARP and DHCP attacks

With ARP and DHCP attacks, the attacker breaks in the process of host initialization in order to inject incorrect host configuration data (for example, default gateway address, DNS server address…) and to establish communication between a host and a rogue device.

Since DHCPv4 request messages are broadcasted, the attacker's rogue DHCP server might send response before the real DHCP server does. This way, the host would receive false connection parameters, including default gateway and DNS server addresses and would thus become vulnerable to man-in-the-middle attacks. By using fake DHCP messages, the attacker might also occupy all of the IP addresses available from the DHCP server's pool.

ARP attacks make use of fake ARP messages, causing incorrect MAC address – IP address entries within the ARP tables. Although the IP address is valid and legal, the victim communicates with the wrong MAC address. The attackers often initiate ARP attacks in order to associate their (attacker's) MAC address with some legal IP address, for example address of the default gateway. Afterwards, they may route all the outbound traffic through their infrastructure, intercept and read the traffic etc.

There are quite a few techniques for ARP and DHCP attack prevention. Cisco's DHCP snooping makes sure that only specific IP addresses with specific MAC addresses on specific ports may access the network. Some IDS-es and other tools can detect ARP spoofing: *arpwatch, anti-arpspoof, ArpON…* Checking on duplicate MAC addresses also helps detecting the attack, while it may also be used for other reasons.

In IPv6, there is no ARP protocol. It is substituted with ICMPv6 protocol elements, namely neighbor discovery protocol. However, the level of security is the same as with ARP messages in IPv4 unless it is used in combination with IPSec. Also ICMPv6 messages may be forged and this leads to the same set of problems as with ARP protocol.

Broadcast and Flooding

Too many messages targeted at the same host or network connection results in interrupted operation – DoS or Denial of Service attack. An echo request is sent to the target broadcast network address, while the source address is spoofed and represents an IP address of the target. All of the network hosts will respond with sending the echo reply and thus flood the target.

In IPv4, most routers enable "no IP directed broadcasts" (Cisco) or similar command to prevent from flooding responses, generated after a broadcasted request, to a single unicast address.

In IPv6, the concept of broadcast is substituted with multicast. Use of ICMPv6 messages, generated as a response to a packet with multicast target address, is generally not advisable or even prohibited since they represent an attack surface. Strict filtering of packets with the multicast address as source address is recommended when there are no real reasons to send traffic from a multicast address.

Routing Attacks

Routing attacks aim at interrupting or rerouting the traffic flows in the network by means of flooding network devices, deleting routes, and sending out fake route advertisements.

In IPv4, routing information is usually secured by authentication, making use of MD5 hashing algorithm. In IPv6, some of the protocols are unchanged while the others are considerably different from IPv4 protocols: BGP and IS-IS are changed and adapted for IPv6. OSPFv3 and RIPng have lost the authentication part: they use AH and ESP headers for the purposes of integrity, authentication and privacy .

Countermeasures against routing attacks are more or less standard: administrators should use the traditional

authentication with BGP and IS-IS, and IPSec to secure OSPFv3 and RIPng.

Viruses and Worms

Viruses and worms are among the most widespread problems in the today's IP networks. In IPv4, they may harm a lot of hosts at the same time, while they are also responsible for increased network traffic. Timely security patches, antivirus software and early detection are three most important defense techniques.

In IPv6, a traditional virus shape is not changed. E-mail viruses and viruses infecting the removable media are the same as in v4. However viruses do have more work to do due to a high number of network addresses that need to be searched in order to find vulnerable hosts. The famous SQL slammer, causing massive damage to the internet, would cause much less trouble in IPv6network!

CONCLUSIONS

In the paper, we have compared several types of well known network attacks in IPv4 and IPv6, motivation behind them and also the most important countermeasures for attack detection and/or prevention. Although here are some attacks typical for either IPv4 or IPv6, we have showed that most of the attacks have similar anatomy. However the similarity should not deceive the administrator: even slight difference in network administration might yield a huge difference between secure and insecure system!

REFERENCES:

Arkko J., Kemf J., Nikander P., Zill B., RFC 3971 – Secure Neighbor Discovery

Arkko J., Aura T., et al.: Securing IPv6 Neighbour and Router Discovery, http://research.microsoft.com/users/tuomaura/publications/arkko+Wise02.pdf

Bound J, Carney M., et al.: RFC 3315 – Dynamic Host Configuration Protocol for IPv6.

Bhaiji Y., Layer 2 Attacks & Mitigation Techniques, Sanog8, Karachi, August 2006. http://www.sanog.org/resources/sanog8/sanog8-L2-attack-mitigation-technique-yusuf.pdf

Bruschi D., Ornaghi A., Rosti E.: S-ARP: a Secure Address Resolution Protocol, pp.66, 19th Annual Computer Security Applications Conference (ACSAC '03), 2003.

Convery S., Miller D.: IPv6 and IPv4 Threat Comparison and Best-Practice Evaluation, Maj 2004. TIndustries Security Whitepaper - http://www.stindustries.net.

Deering S., Hinden R., RFC 2460 – Internet Protocol, Ver.6, specification. http://rfc.sunsite.dk/rfc/rfc2460.html

Dorms R., RFC 2131 – Dynamic Host Configuration Protocol.

Hines. A. Neighbour Discovery in IPv6, University of Paderborn, http://wwwcs.upb.de/cs/ag-madh/WWW/Teaching/2004SS/AlgInternet/Submissions/17-neighbour-discovery-protocol-in-IPv6.pdf

Kempf J., Nikander P., Nordmark E., RFC 3756 – Ipv6 Neighbour Discovery Trust Models and Threats.

Marn B.: Attacks on IPv4 and IPv6, B.Sc. thesis, Faculty of Computer and Information Science, University of Ljubljana. 2008.

Narden T., Nordmark E., Simpson W.: RFC2461: Neighbour Discovery for IP Version 6.

Ptacek T.H., Newsham T.N., Simpson H.J.: Insertion, evasion, and denial of service: Eluding network intrusion detection. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.5765.

Sušnik R.: IPSec Protocol in IP Networks, B.Sc. thesis, Faculty of Computer and Information Science, University of Ljubljana. 2001.

Tremblay J.F.: Introduction to ICMPv6 Neighbor Discovery, Autoconfiguration and IPv6 DNS, www.hexago.com/4105/file.asp?file_id=395

Van Hauser, Attacking the IPv6 Protocol Suite, http://pacsec.jp/psj05-vanhauser-en.pdf.

Whalen S., An Introduction to ARP Spoofing. http://www.rootsecure.net/content/downloads/pdf/arp_spoofing_intro.pdf

AUTHOR BIOGRAPHY

MOJCA CIGLARIČ received her PhD in computer science in 2003. Her research covers the area of distributed systems. Currently she works at University of Ljubljana, Faculty of Computer and Information Science as an assistant professor.

MATJAŽ PANČUR received his PhD in computer science in 2006. His research area covers agile methodologies and distributed systems. Currently he works at University of Ljubljana, Faculty of Computer and Information Science as a teaching assistant.

ANDREJ KREVL is currently finishing his MSc in computer science. His research covers all areas of networking. Currently he works at University of Ljubljana, Faculty of Computer and Information Science as a teaching assistant.

AUTHOR LISTING

AUTHOR LISTING