Adaptive and personalized multimedia content delivery for disabled users in Internet TV

I. Anagnostopoulos[1], C.N. Anagnostopoulos[2], E. Vlachogiannis[3], D. Gavalas[2], and G. Tsekouras[2]

[1] Department of Information and Communication Systems Engineering, University of the Aegean,

Karlovassi, Samos island, 83200, Greece,

E-mail: janag@.aegean.gr

[2] Department of Cultural Technology and Communication,

University of the Aegean,

Mytilene, Lesvos island, 81100, Greece,

E-mail: {canag,gtsek}@ct.aegean.gr, dgavalas@aegean.gr

[3] Department of Product and Systems Design Engineering,

University of the Aegean,

Hermoupolis, Syros island, Greece

E-mail: evlach@aegean.gr

Abstract. This paper aims at the delivery of adaptive and personalized multimedia content in interactive IPTV environments, using programmable IP services through the MPEG-21 standard, supporting also features for users with disabilities. Moreover, we propose a system that adapts to users' preferences using profile separation, not only for individual users but also for user groups as a whole. The system takes advantage of explicit and implicit information through the users' interaction with the IPTV environment, while the profile reflects groups of similar users, thus dropping the time needed for matching user patterns and profiles when forming a recommendation. The system works in conjunction with a simulation platform acting as an interaction interface between the IPTV architecture and the prospective viewer. Based on that, interactivity in IPTV is faced through metadata and adaptation.

Keywords: IPTV, user interaction, personalization, adaptation

1 Introduction

Nowadays, the amount of available disseminated digital content is increasing over the Internet and broadcast network technologies. Thus, more and more augmented personalized schemas and mechanisms are required by the content consumers (users), who desire to adequately deal with this vast amount of information. Personalization increases the possibility of creating a plethora of new services for both users and content providers. Thus, a provider may broadcast the same multimedia content to the consumers, while their front-end terminals display differentiated interactive services, according to the user profile such as personal preferences, location-based characteristics, users' disabilities etc.

In order to create personalized multimedia content, appropriate media elements and metadata must be selected from the multimedia storages on the basis of the user profile information and contextual situation. Metadata, also called data about data, provide enhanced content description including characteristics about the data such as accuracy, reliability and other qualitative features. Metadata provide the mechanism to describe data in a consistent form allowing users to gain a uniform understanding of the content and purpose of datasets.

Most metadata formats are XML-based, thus having hierarchical structures suitable for most types of content format (MPEG-7 [1] and MPEG-21 [2]). For instance, information can be displayed on the client screen in the interactive IPTV paradigm, while playing the audiovisual content. Apart from metadata, rich media is also a key technology, especially for multimedia content and interactive services, enhancing the synergy between the traditional broadcast networks and Internet technologies [3].

In addition, IPTV is now considered to be one of the key growth areas for telecom/broadcast operators, as well as content providers. Related technologies that use the Internet Protocol infrastructure, are continuously emerging globally, with promising user-centric prospects, while interactive services allow the user to not only receive information but to control and/or process it. Another interesting service in IPTV environments are recommendation systems. These systems provide personalized TV models based on user modelling and browsing behaviour. The term browsing behaviour is used in order to highlight the interactivity of the user similarly to the interaction the user has when browses the web. TV recommender systems have the potential to become essential tools not only in aiding people to choose what content they will consume but also how to consume the same content if they have disabilities [4], [5], [6], [7], [8], [9]. In order for these systems to work properly and achieve their aims there are many problems and issues that must be solved from other research fields such as Human Computer Interaction, Psychology etc.

In this paper a novel approach for intelligent content personalization in Internet TV using the new MPEG-21 standard is introduced aiming at the delivery of interactive services to users with disabilities. Moreover, we propose a recommendation system that adapts to users' needs using profile information either from a single user or from a specific user group as a whole. The system takes advantage of explicit and implicit information through the users' interaction with the IPTV system and the profile reflects groups of similar users, thus decreasing the time needed for matching user patterns and profiles when forming a recommendation.

The remainder of the paper is organised as follows. In the next section, a survey of the related work in the field of recommendation systems for interactive TV environments is presented. Section 3 presents an overview of the features of IPTV as well as of the MPEG-21 multimedia framework. In Section 4 we describe our system, which performs personalization and profile separation. Finally, Section 5 presents all design issues and implemented sub-systems regarding the IPTV simulation platform we used, while Section 6 concludes and summarizes our work.

2 Related work

This section describes the related work in the field of recommendation systems for interactive TV environments. Some approaches use implicit, while others explicit techniques for processing user browsing behaviour and personalized preferences. In [10] the authors propose an Internet-based recommendation system named PTV. Through this system users register, login, and consume their personalized recommendations as specially customized HTML. The system produces personalized TV guides by integrating user profiling, content-based reasoning, and collaborative profiling techniques. The system also supports WML for mobile phones and PDAs. However, in this approach only explicit information is used, meaning that no implicit information is used in order to extract users' interaction history or behaviour.

Other approaches are actually belong to the recommendation type of systems called Electronic Program Guides (EPGs) [11], [12]. These systems use multi-agent architectures, where specialized software modules (agents) collect data about available events and monitor the user's interaction behaviour, through a multimodal user interface.

The authors in [11] and [12] propose a Personalized EPG that employs an agent-based system designed for set-top box operation, by using three user modelling modules, which collaborate in preparing the final recommendations. The modules used explicit, stereotypical and dynamic user preferences.

Moreover, in the works described in [13] and [14], multi-agent TV recommendation systems are designed to work together and collectively model the profile information on a particular user. The systems both include implicit and explicit methods since they encapsulate implicit interaction history, explicit TV viewing preferences, and feedback information either implicitly or explicitly conducted.

To be more specific, the authors in [14] proposed a hybrid multi-agent recommendation system, which combines one explicit recommender agent with two implicit recommender agents through the implementation of Bayesian methods and Decision Trees. In addition, the authors in [15] developed the TV Advisor, where users enter their explicit preferences in order to produce a list of recommendations.

Current EPGs found in products such as DirecTV and EchoStar's digital satellite set-top boxes, cable set-top boxes from Time-Warner Cable and Cable-Vision, and personal video recorders by TiVo and ReplayTV, offer users several methods for searching and browsing TV listings [16].

3 Internet Protocol Television and MPEG-21

3.1 IPTV – an overview

One of the most promising services in Internet technology is without doubt IPTV (Internet Protocol television). In this technology, television and/or video signals are distributed to subscribers or viewers using a broadband connection over the Internet Protocol using packets over a network infrastructure. The distinctive feature of this service is that interactivity can be easily provided, since the user can compile the program according to his individual preferences. Apart from classic programs from broadcasters and TV stations, the content can also be watched with a time delay. IPTV networks work similarly to cable systems whereby video/audio signals are groomed at a centralized head end and distributed over a private network. With IPTV, the customer base employs a two-way, digital equipment, allowing for interactivity.

However, sometimes IPTV is often provided in conjunction with Video on Demand and may be bundled with Internet services such as Internet access and VoIP. The physical path that delivers the service to users does not always come to the user over the Internet, but over carrier owned fiber optics, or coaxial cables.

The benefits arising from IPTV have not only one dimension, such as interactivity, but many others arise as well for both users and content providers. The user can receive telephone, data and video for a monthly cost through an Internet provider undercutting combined costs of getting those services delivered by other means. On the other hand the provider may achieve the exclusivity of content for a specific target group of viewers. The provider can also increase its profits, through a variety of extra services such as billed personalized content delivery and/or targeted ads. Finally, IPTV as an IP-based service has the advantage of integration and convergence, mostly when using IMS-based solutions [17]. Converged services imply the interaction of existing services in a seamless manner to create new value added services.

As far as the limitations in IPTV are concerned, most of them are due to the fact that IPTV design requirements are fragmented. Suppliers and providers are still struggling to find sizeable commercial deployments to which they can sell their products in volume (information found at http://www.gdbtv.com/what.html). IPTV technology has also the drawback of different regional regulations and law frameworks regarding the physical layer and the way the provider uses cable, fiber, satellite or terrestrial means. Available bandwidth and data rates also vary among different Internet providers, while there is no standardisation among requirements for conditional access and digital-rights management. Also, issues such as integrating different management policies such as billing services and similar consideration are still unsolved. Finally, since IPTV requires real-time data transmission by utilising the Internet Protocol, it is sensitive to packet loss and delays in those cases where network services are unreliable and mostly when streaming multimedia content across wireless links.

3.2 The Multimedia Framework MPEG-21

MPEG-21 Multimedia Framework is based mainly on two essential concepts: the definition of a fundamental unit of distribution and transaction named as the Digital Item, as well as the concept of users interacting with these units. Digital Items can be considered as the object of the Multimedia Framework (e.g., a video collection, music album), while users can be considered as the objects of the Multimedia Framework.

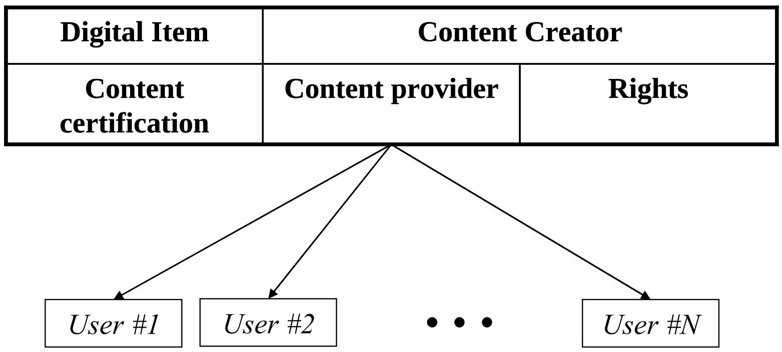


Figure 1. Layers and entities in MPEG-21

According to the technical report described in [2] the goal of MPEG-21 is to define the technology needed to support users to exchange, access, consume, trade and/or manipulate Digital Items in an efficient, transparent and interoperable way. In particular, this report gives an overview of the technologies that have been identified to enable this goal and that are consequently being reflected in the different parts of the MPEG-21 standard. The main entities involved in MPEG-21 are users and Digital Items as illustrated in Figure 1.

Moreover, an MPEG-21 User is any entity that interacts within the multimedia framework, thus it can be a content creator, a content distributor, or an end user. End users include individuals, communities, organisations, corporations, or consortia. Users are identified specifically by their relationships to others users for a certain interaction, and thus MPEG-21 makes no distinction between a "content provider" and a "consumer" since both are treated as Users in the Multimedia Framework.

The Digital Item is the basic unit regarding distribution and transaction among different users in the framework. It is a structured digital object with resources (e.g. videos, images, documents) holding a unique identification and corresponding metadata structured in XML schemas. The structure corresponds to the relationships among the parts of the Digital Item, which are both resources and metadata. Once the content (in the form of Digital Items resources) will be exchanged in the defined framework, there will be entities that will offer content customization functionalities to achieve an optimal end user experience. Therefore, MPEG-21 as a whole sets the trail to create a complete system, where specific entities play the role of the "bridge" between parts that have to be somehow related, such as the multimedia content and the interface environment [18].

4 The proposed system

4.1 Overview of the architecture

As previously mentioned MPEG-21 is considered as the main medium for integrating rich interactive services into IPTV. Based on that, we planned a full scenario that includes production, delivery and use of the disseminated digital content, aiming at identifying the primary entities and the way these are involved in the overall system design. Thus, as depicted in Figure 2, the Content Designer (ConDes) identifies and describes the target groups using their profile (people with disabilities) and/or their preferences associating in parallel interaction modes (e.g. auditory description for visually impaired people, subtitles for hearing-impaired etc). By using authoring tools Content Designers develop the required content components upon the selection/adaptation on the interaction modes.

In case where a hearing-impaired end user wants to consume developed content the context of use is accompanied with attributes like access device capabilities, text appearance, time and location of the end user. The delivery of this context is accompanied upon the user request as well as according to the already provided profile. Then, the system maps the user's context of use with an appropriate composition of the components of the content. If, during content consumption, the context of use is being modified, the system needs to be aware so that it can adapt to new requirements.

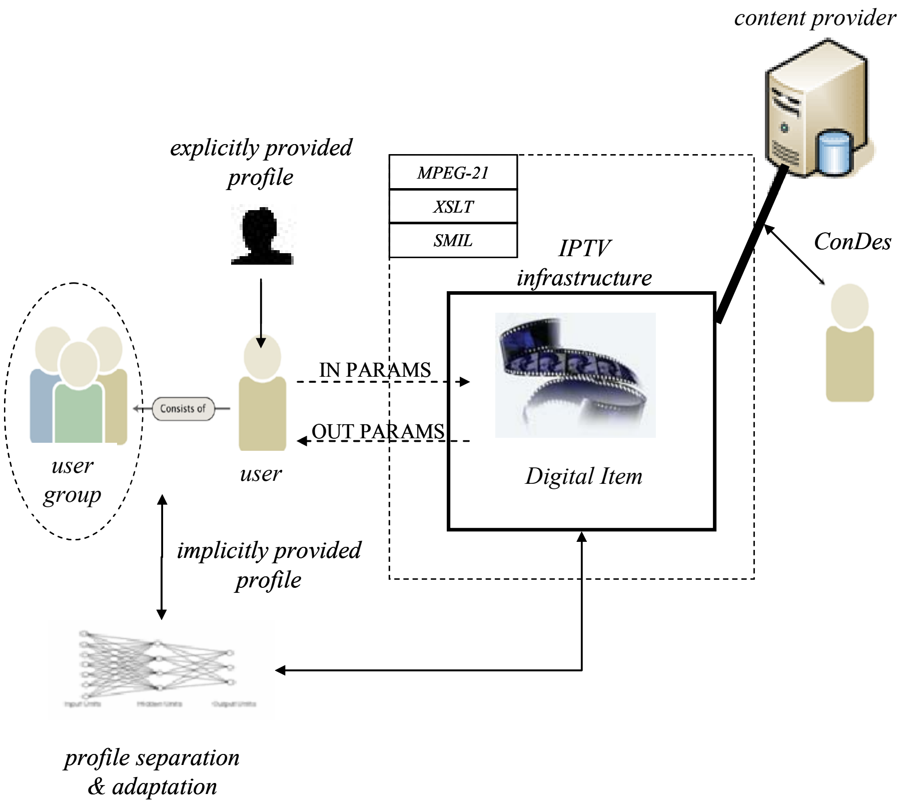


Figure 2. Actors and related technologies in the proposed system

Even if MPEG-21 addresses considerations for adaptation and specifically accessibility by providing several related description elements into its schema, this cannot ensure the accessibility of delivered content. However, this is a fundamental condition for providing accessibility output of the systems involved. In other words, it should be able to provide the required infrastructure so that a digital content would be able to obtain the necessary variety for both the content designer, to be able to design accessible content, and the involved systems, to have the required information to deliver an accessible result. Figure 3 presents the stakeholders from the adaptation point of view, where the content provider, the author (also referred to as content designer), the authoring tools, the systems of the content provider and of course the consumer with his associated interaction profile (i.e., preferences, device capabilities) are identified and all play a major role to the iTV accessibility.

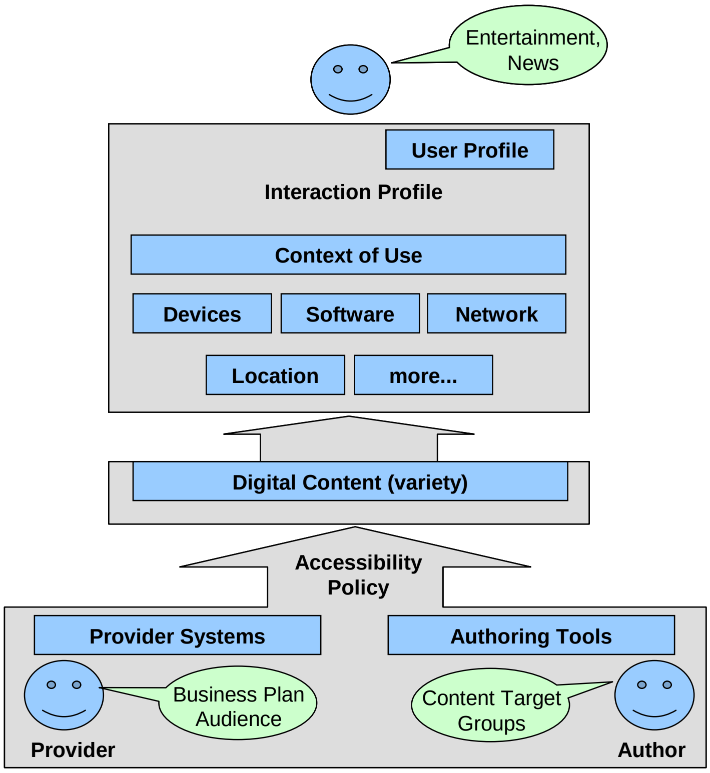


Figure 3. Multimedia delivery stakeholders

To sum up with, the role of the MPEG-21 towards accessibility is revealed through the following aspects:

- *Alternative content:* MPEG-21 offers metadata (as defined in MPEG21 - Part 7 [19]) that allows content providers to provide the content in one or more alternative ways. The ways often refer to different modalities and thus they can include captions, audio descriptions, etc.

- *Digital content navigation:* In interactive IPTV environments, navigation facilities within available content are provided by an Electronic Program Guide (EPG). This is actually the interactive portion of the system that offers the required functionality to the user including service selection and retrieval, service information and scheduling, updating profiles and personal preferences, rating processes and mechanisms etc.

- *Description of context of use (IN PARAMS):* The usage context actually refers to all the information that needs to be taken into account to adapt digital content according to the user's requirements.

- *Description of presentation parameters of digital content (OUT PARAMS):* This determines what technical characteristics need to be adapted. An important implementation consideration was the transformation of MPEG-21 to the

Synchronized Multimedia Integration Language - SMIL format as an intermediate solution to ensure media players' compatibility. This involved the mapping between those two schemas realized using XSLT. SMIL, is a W3C recommended XML markup language and defines markup for timing, layout, animations, visual transitions, and media embedding. It allows the presentation of multimedia items such as text, images, video, and audio, as well as links to other SMIL entities and files from multiple Web servers.

- *Device accessibility:* This refers to the accessibility of the involved hardware including remote controls and set-top boxes

- *Content provider accessibility policy:* An important contribution to accessibility issues in MPEG-21 is the capability of applying and claiming for an accessibility policy. In other words, content providers need to be capable of applying a kind of accessibility policy based on the target consumer group and the former requirements for quality assurance. For instance, such a policy could require digital content to be accompanied by subtitles of two languages (e.g. English, Greek) and every image by an alternative text between two and ten words. Applying such policies requires a mechanism for validating a digital content against a policy description and could be for instance implemented based on Schematron (http://xml.ascc.net/schematron/schematron1-5.sch), which is an XML structure validation language for making assertions about the presence or absence of patterns in trees.

4.2 Personalization and profile separation

In our proposal the binding of metadata and XML descriptors with the actual multimedia content (i.e. video) is performed, thus creating the integrated Digital Item. A content manager is capable of controlling, refreshing and synchronizing metadata files in conformance with their content including detection and/or elimination of XML metadata (sub-)item (or component) duplicates.

When a Digital Item is processed, it is in the appropriate form to interact with an intelligent software module, which is able to assign a user/viewer to a specific category (class) matching the content that best suits to his preferences. The module that performs this process is a three-layer probabilistic neural network, which is properly trained and capable of classifying user profiles according to a selected taxonomy of profiles. User preferences and activities (browsing behaviour etc.) are implicitly gathered through the user's interaction with the system itself. The architecture that performs content personalization is illustrated in Figure 4. On one hand, personalization augments interaction properties, allowing users to browse programs much more efficiently according to their preferences. On the other hand, it also enables to build social networks that can improve the performance of current IPTV systems considerably by increasing content availability, trust and the realization of proper incentives to exchange content.

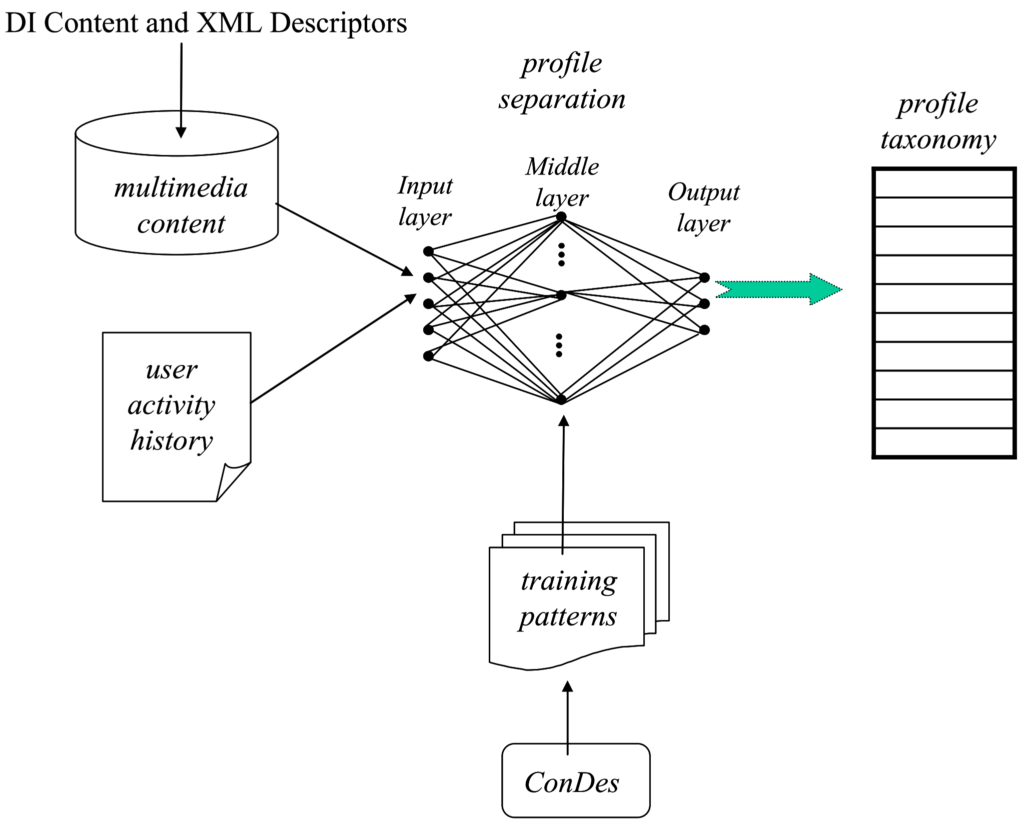


Figure 4. Profile separation joined with an artificial intelligence system for content personalization

Profile separation is achieved through an artificial intelligence classifier, examining content and XML descriptors. The software module that performs this action is a three-layer Probabilistic Neural Network (PNN). This type of neural networks combines some of the best attributes of statistical pattern recognition and feed-forward neural networks. PNNs are the neural network implementation of kernel discriminate analysis and were introduced into the neural network literature by Donald Specht [20]. A useful characteristic of such a kind of neural networks is that they present very short training times, which makes them suitable for real-time systems, in contrast to larger memory requirements and slower execution speed for prediction of unknown patterns compared to other types of neural networks [21]. A probabilistic neural network uses a supervised training set to develop distribution functions within a pattern layer. These functions, in the recall mode, are used to estimate the likelihood of an input feature vector being part of a learned category, or class. The learned patterns can also be combined or weighted, with the a priori probability of each category, to determine the most likely class for a given input vector. If the relative frequency of the categories is unknown, then all categories can be assumed to be equally possible and the determination of the category is solely based on the closeness of the input feature vector to the distribution function of a class.

Probabilistic Neural Networks contain an input layer, with as many elements as there are separable parameters needed to describe the objects to be classified as well as a middle/pattern layer, which organises the training set so that an individual processing element represents each input vector. Despite the fact that PNN and other radial-base neural networks actually suffer more from the "curse of dimensionality" than other linearly-based neural networks, if the number of the input nodes is properly chosen, then the specific problem would not appear [21]. We particularly took under consideration that issue, by satisfying that the ratio of the pattern versus the input layer should be greater or equal to ten. Finally, they have an output layer also called summation layer, which has as many processing elements as there are classes to be recognised. Each element in this layer is combined via processing elements within the pattern layer, which relate to the same class and prepares that category for output.

A numerical example regarding the proposed PNN architecture is 8-M-O, where $M \geq 80$ and O equals to the number of predicted classes. In other words, this means that when we apply a PNN with 8 input nodes as the responsible classification features, we will avoid the "curse of dimensionality" if we use relative information from at least 80 content consumers. Table 1, depicts an instance of eight features chosen as the input nodes, where two are implicitly gathered features and the rest are explicitly given by the user. XML data

<timestamp>, <action>, <type>, <content rating> will be described in the following sections.

INPUT NODES	
IMPLICIT DATA	**EXPLICIT DATA**
XML data <timestamp> *XML data <action>*	*XML data <type>* *XML data <content rating>* *IN PARAMS (usage context)* *OUT PARAMS (layout)* *OUT PARAMS (visual – audio context)* *Content Provider*

Table 1. Data in the input nodes

5 The IPTV Simulation Platform

In the framework of a Greek National project a simulation platform was developed, acting as an interaction interface between a iTV architecture and the prospective viewer. A user interface prototype has been implemented to enable users to effectively browse, search, download and consume the provided multimedia content. In the case of disabled people, "effectively" means that both the content and all related accessibility services need to be accessible by the user. Such an interface is actually a sub-system of the overall system architecture as briefly presented in Figure 5, consisting of Content Descriptors created by an authoring tool, storage procedures (native XML database) and a software module that performs profile separation. As described in the previous section, the software module that performs profile separation is an artificial neural network capable of classifying digital items and user profiles based on their attributes and the user's browsing behavior, enable thus program recommendations. All procedures are performed through web services and under a flexible and distributed architecture.

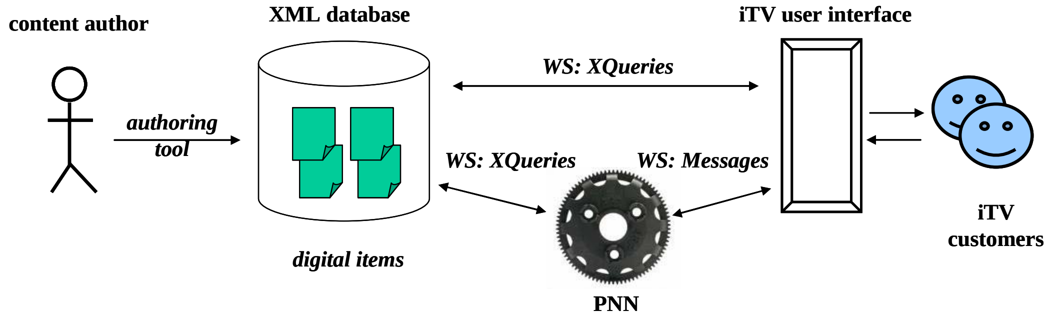


Figure 5. iTV adaptation architecture

5.1 Design Issues

This section describes a user interface that involves an Electronic Program Guide (EPG) simulator. It should be noted that the choice of the implementation technologies has not been straight-forward considering the plethora of available standards and technologies like MHP4, GEM-IPTV, TV-Anytime, DVB-IP, Java-TV and more. Given the requirement for incorporating networking functionality into the EPG subsystem, a Web-based approach instead of a standalone application has been adopted. This approach ensures execution of the EPG through a standard browser interface, where the main actors are the End User, the Service Provider and the Program Recommender.

The End User interacts with the interface browsing according to his personal preferences. The user's browsing behaviour is associated with XML-based features. These XML data hold user profile information which includes personal data such as potential disabilities (hearing problems, visual impairments, etc), preferences upon the audiovisual content (e.g. sports, news, movies), browsing behaviour (interaction time intervals, navigational patterns etc.). On the other hand the Service Provider serves the End Users, informing them about the offered services as well as their availability time schedule.

Finally, the Program Recommender acts according to the decisions made by the profile separation module and the XML-based feature created implicitly by the user's browsing behaviour. In the following section we analyze the functional and interactivity requirements of the above-mentioned subsystems discussing also the solutions adopted in our prototype.

5.2 Implemented sub-systems

The implemented sub-systems are the Initiator, the Adaptation engine, the Player, and the EPG. The first one is the software module that initiates all necessary procedures, providing also feedback to and from the user, either implicitly or explicitly through the Adaptation engine. Both modules record the users' actions (e.g. profile modification, starting/pausing/resuming a TV program, history records etc).

5.2.1 The Adaptation Engine

The Adaptation Engine uses a personalization algorithm that weights the relevance of the consumed content, based on the user interaction with the system. In our approach, we assume that past interaction behaviour is an indicator of the user's future interaction with the system. The construction of the personalized browsing behaviour is performed in a totally transparent way, ensuring in this way no biases. The personalized patterns are recorded and updated continuously according to the content consumed by the user (e.g. sports, news, kids movie etc), the time spent until his next interaction with the system as well according to his navigational patterns, his history and the content he consumed upon starting the Initiator. The main target is to continuously adjust the user's profile to any possible changes in respect to his navigational patterns. The stored data define behavioural patterns regarding the information explored by the user through the content consumed.

Thus, from the beginning of the session (the time the Initiator starts) we record six features, namely the session and the user IDs, the Content Provider, the type of the content consumed, the Timestamps between subsequent browsing activities and interactions as well as the actions the user made.

Every time the user consumes multimedia content, these features are written into an XML file updating in parallel the weights that define the user's confidence with respect to each of the consumed content, according to the mechanism described in the following. All these data are implicitly and explicitly gathered through the user's interaction with the system.

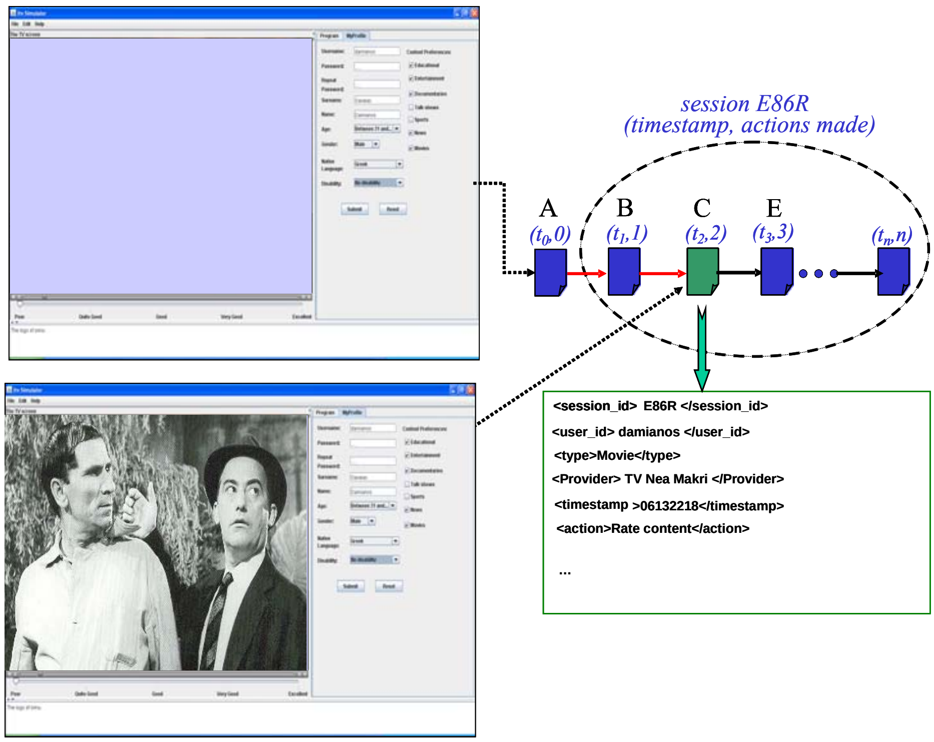


Figure 6. Recording interaction features

An example of the recorded XML instances file is presented in Figure 6 where a specific user (logged-on as ‘damianos’) inserted explicitly his preferences in the Initiator and started to consume content (interaction instance: A, timestamp t_0). The user then decided to watch news acquired from a specific content provider (interaction instance: B, timestamp t_1) and then switched the content to a Greek movie (interaction instance: C, timestamp t_2) where he actually intended to rate the content. In Figure 6, values that equal from 0 up to *n* inside the parentheses, represent different user actions (value 0 represents the log-on procedure) and are enumerated in this way just for simplicity reasons.

The XML record regarding interaction instance C is illustrated in the right part of Figure 6. This instance records the session <*session_id*>, the user <*user_id*>, the content type <*type*>, the Content Provider <*provider*>, the day and the time of a specific interaction instance <*timestamp*>, and finally the actions made from the user while consuming the content <*action*>. Note that in a recorded instance more than one <*action*> tags may appear.

The timestamp for each instance is defined by using a function similar to the PHP time() function, which is the Unix timestamp (the number of seconds since January 1 1970 00:00:00 GMT). In our approach, the time spent between subsequent interactions with the system is calculated by the difference between subsequent recorded timestamps. Thus, in our example, the time the user spent to watch news before switching his preferences to a movie is (t_2-t_1).

Now in order to personalize the similarity of a content consumed by a user with respect to his browsing behaviour, we used two probabilistic functions. These functions assign a probability value according to the time intervals between subsequent interactions where the user consumes content for information exploration as well as according to the number of actions made by the user until he reached the desired content. We consider that time is an important factor since the more time the user spends consuming a specific content, the more this content is possible to be relevant to his preferences.

Moreover, the time spent during a search session is modelled with a lognormal distribution. This kind of distribution was selected among others, given that this distribution fits with the results made in similar investigations (information found at http//www.mondosoft.com/SearchBehaviorWP.pdf). In this work, it was concluded that users likely desire to consume multimedia content fast and with the minimum possible effort. In other words, the user wants to avoid as much as possible to interact with the system, while he prefers to have recommendations or suggestions. Too much interactivity may end in a non user-friendly environment and may confuse the user. Since a typical behaviour of a user looks a lot like to the aforementioned pattern, we modelled the user's interaction behaviour by a lognormal distribution function, where σ=1 and θ=0.5 as depicted in Figure 7(a).

According to this model, if a user presents a rather small grade of interactivity when consuming content for a significant time interval then this user is more focused on his needs and more reliable when automatically measure his preferences, in comparison with another user who ends to consume the same content after a lot of interactivity.

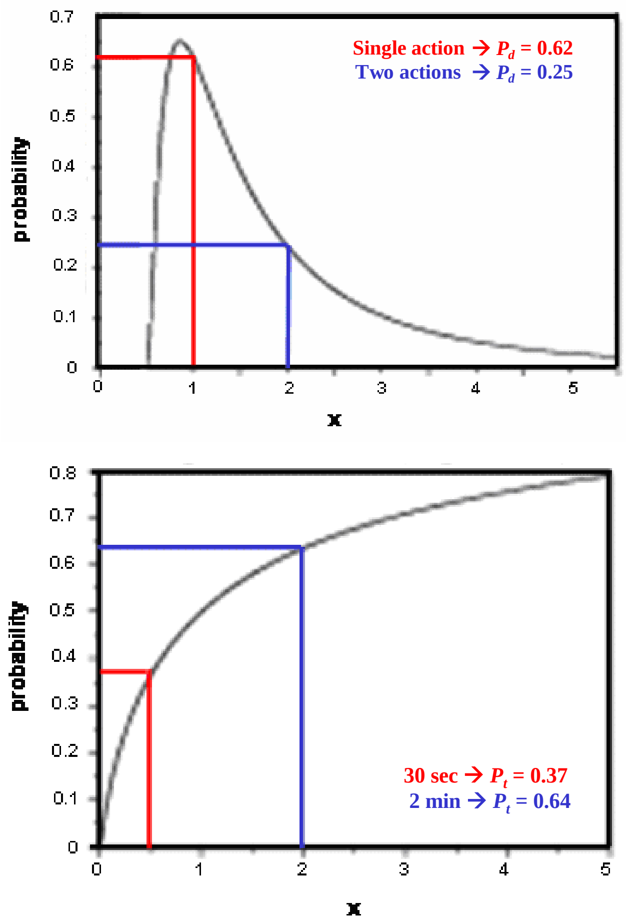


Figure 7. Distributions (lognornal/cumulative) for modeling user interaction characteristics

As far as the time spent between subsequent interactions is concerned, we again relied on the behavioural patterns described in (information found at http//www.mondosoft.com/SearchBehaviorWP.pdf). However, since time is a crucial parameter and in order to avoid biases when recording the user profile we assumed that if a user consumes a specific content for at least five minutes before interacting again with the system the selected content is highly relevant to his profile. We had to set such a threshold in order to exclude cases where the user has some intervals of inactivity and idle periods (e.g. the user has a telephone call, or he leaves from the room or gets some sleep etc.)

Thus the time spent during subsequent intervals, was modelled with a lognormal cumulative distribution function, where σ=2, for the interval that spans between zero and five minutes. Figure 7(b) shows an example, where according to this model, if a user spends 30 seconds for consuming a specific content and then interacts again with the system (either terminates the session), this content is expected to be 37% relevant in respect to his profile, while when he spends 2 minutes it is considered as 64% relevant.

As a session set, we define all recorded XML instances, in which the user interacts with the system after his registration in the Initiator as illustrated in Figure 6. If the user interacts *N* times with the system then *N* XML instances are recorded.

Equation 1 defines the lognormal distribution probability density function where θ=0.5 and m=σ=1, while Equation 2 expresses the lognormal cumulative distribution function, where Φ is the cumulative distribution function of the normal distribution and σ=2. Eventually, based on these equations, we define the consumer's browsing behaviour score (*Consumer Score - ConS*) according to Equation 3, which actually defines the maximum normalised product between P_t and P_d. The reason for taking under consideration only the maximum value and not all the rest values is because in this way we noticed that we can discriminate users with different content preferences. For example a user who is particularly interested in athletics, will have a larger *ConS* value when he watches a sports game rather when he consumes other multimedia content. Parameters c_1, c_2 are used for fine-tuning purposes and their values are positive. This allows to weigh the *ConS(t,d)*, according to the time intervals *t* between subsequent interactions as well as to the amount of actions made *d* until the content consumption. Fine-tuning of c_1 and c_2 is quite important and gives the opportunity for further adjustments over individual user behaviour and over parameters *t* and *d*. However, in this paper we consider that both behaviour metrics contribute equally to the personalized score and thus the fine-tuning parameters are set equal to one.

$$P_d(x) = \frac{e^{-((\ln((x-\theta/m))^2/(2\sigma^2))}}{(x-\theta)\sigma\sqrt{2}\pi} \qquad (1)$$

$$P_t(x) = \Phi(\frac{\ln(x)}{\sigma}) \qquad (2)$$

$$ConS(t,d) = \frac{\max[P_t^{c_1} \cdot P_d^{c_2}]}{\sum_{d=1}^{l} \int_{t=0}^{T} (P_t^{c_1} \cdot P_d^{c_2})dt} \qquad (3)$$

5.2.2 The Player subsystem

This system reproduces Digital Items recording also the user's interaction history. The basic module used is the digital content player, which supports not only simple functionalities (play, pause, rewind, etc.), but enhanced ones, such as subtitles, audio descriptions, etc. In our work, we used the Synchronized Multimedia Integration Language (SMIL) as intermediate technology mainly since MPEG-21 Digital Item Declarations can be easily transformed to SMIL format by using XSLT transformation, ensuring thus a user-friendly and interoperable environment through open Internet/Web technologies. In our approach, the SMIL player was implemented using the QuickTime for Java API (http://developer.apple.com/quicktime/qtjava).

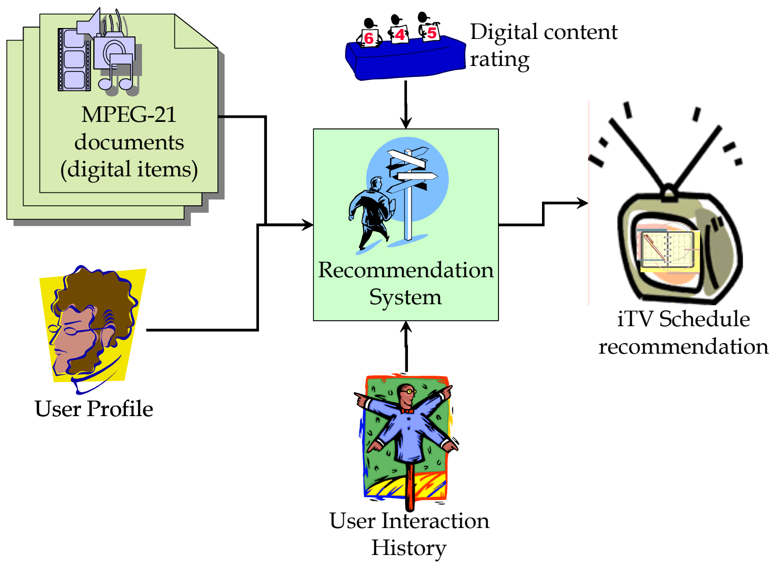


Figure 8. User interaction and recommendation

The XSLT transformation of MPEG-21 Digital Items to SMIL documents depends on the user profile, taking also into account potential user disabilities, which are stated explicitly by the user in the Initiator. The second function of the Player subsystem is the provisioning of user interaction information to the profile separation software module. As mentioned before, an XML-based description of the user interaction is first recorded in XML-based files on the client-side and retrieved by the profile separation module to enable more effective and reliable reasoning. As shown in Figure 8, the users' interaction history and the TV programs ratings posted by users that belong to the same target group of content consumers (e.g. people with same disabilities) constitute the input of the recommendation system.

5.2.3 Electronic Program Guide

The Electronic Program Guide (EPG) provides interaction features to the user. By using EPG the user browses, navigates and consumes audiovisual content, while it provides the interface through which an end user may personalize the audiovisual content based on his potential disabilities and/or content preferences as well as schedule a reminder for an IPTV multimedia content.

An important consideration task during the EPG development has been the representation and retrieval of the TV schedule. To satisfy this design requirement we have used TV-Anytime Programme metadata (information downloaded from http://www.bbc.co.uk/opensource/projects/tv_anytime_api/).

The overall functionality of the EPG has been based upon the specifications of the JAVA TV API (JSR-000927).

6 Evaluation

According to the literature, the evaluation procedure is divided in five generic categories namely learnability, memorability, efficiency, effectiveness and user satisfaction [22]. However, in our approach we performed experiments in order to evaluate the usability of the interactive platform in terms of efficiency, effectiveness and learnability, mainly with kids and people with visual and hearing disabilities. Moreover, we defined efficiency as the ratio of the correctly performed user tasks, effectiveness as the time needed for the completion of the user task versus a specific threshold value (depending on the user task). Finally, we defined learnability as the improvement or not of the effectiveness due time.

6.1 Experimental procedure

The evaluation target group was consisted of thirty users. Among them, there were ten juvenile users with a specific disability (five were hearing impaired and five visually impaired). There were also ten juvenile users and ten adult users with no disabilities. As far as the gender distribution in our target group is concerned, there were 16 male and 14 female users. We would like to mention in this point that no user had an experience in using an interactive-based TV environment, nor there were familiar in using an EPG.

Evaluation took place in the laboratories of the Department of Cultural Technology and Communications of the University of the Aegean in Mytilene, during March and April of 2008. The evaluation users were asked to fill some questionnaires regarding demographic issues (age, gender, language, potential disability), their watching habits (how often they watch TV, if they watch TV alone or with their friends), their preferences in specific content or category (e.g. sports, science fiction films, kids movies etc). Then, the users used the EPG of our system in order to accomplish the following tasks:

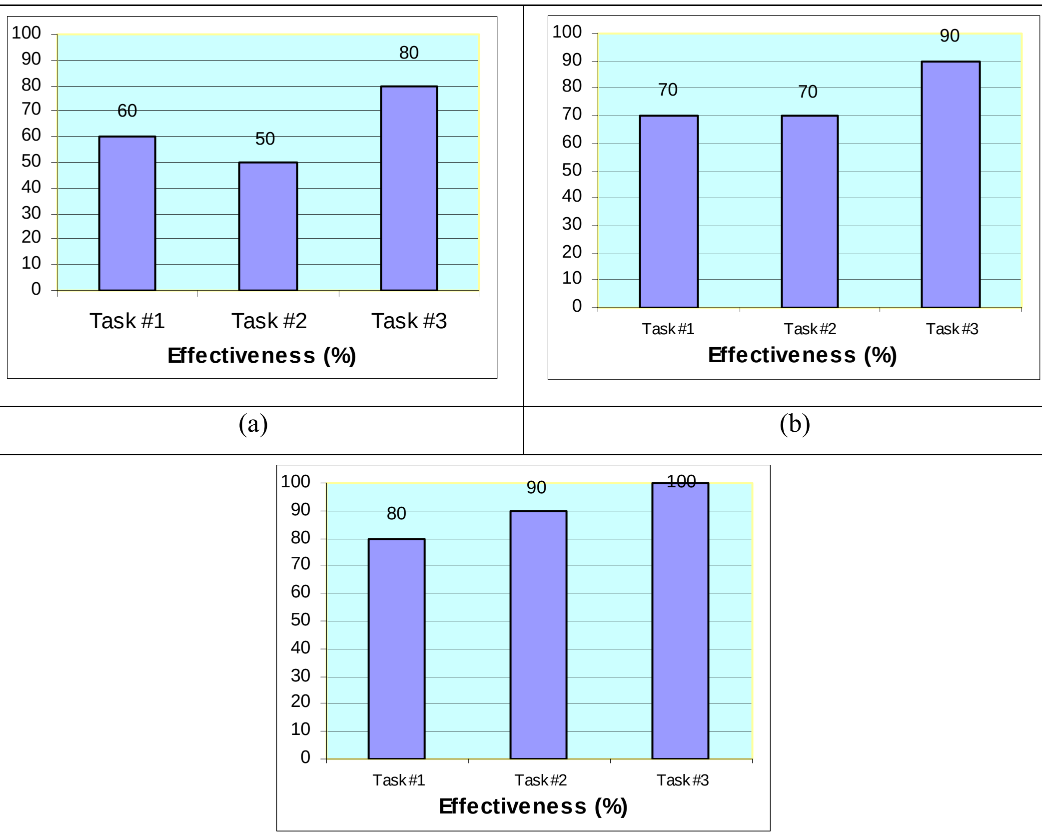
- *Task 1:* Find and reproduce a specific multimedia content,
- *Task 2:* Find and reproduce a specific TV program from several Content Providers,
- *Task 3:* For a predefined time, repeat Tasks 1 and 2 for many TV programs and multimedia content, monitoring potential changes and suggestions proposed by the system through its interaction with the user.

The above-mentioned tasks were repeated twice from each user of the evaluation group. We have to note here that due to the fact that no voice function was supported in the evaluation period, the use of EPG was not fully functional for visually impaired persons. For that reason, a laboratory staff was assigned to read and announce the respective messages from the EPG, helping in that way the respective users to accomplish their tasks.

After the end of the evaluation period and the completion of all tasks, the users were asked to record their experience (e.g. how friendly was the IPTV simulation platform, how easily the user reached the desirable content, which was the most reliable service provided etc). In addition, disable people were asked to evaluate the provided services (e.g. how helpful the voice description was for the visually impaired, how descriptive the written comments were for the hearing impaired person etc). In parallel, after the end of each evaluation session, the completed questionnaires were collected and all users participated in an oral interview. Through this procedure the users expressed the experiences they had through their interaction with the platform, highlighting also potential problems and omissions from our site. We also received a valuable feedback of proposals for improving the services provided, or for introducing new services, augmenting in this way the proposed platform.

6.2 Result analysis

As mentioned in the previous section, we performed experiments in order to evaluate the usability of the interactive platform in terms of efficiency, effectiveness and learnability. Figures 9a, 9b, and 9c depict the effectiveness measured for user groups between 8 and 12 years old, 13 and 17 years old as well as over 18 years old. All results are in respect to the three user tasks. Through the analysis of these results, it was clear that for all users, irrespectively of their age, it was more difficult to accomplish the first two tasks (Task#1 and (Task#2), while it was not so hard to accomplish the last task (Task#3).



(a) (b) (c)

Figure 9. Effectiveness measurement for users between (a) 8-12 years old, (b) 13-17 years old, (c) over 18 years old.

Similarly, Figures 10a, 10b, and 10c, illustrate the efficiency measured for user groups between 8 and 12 years old, 13 and 17 years old as well as over 18 years old, in respect to the three user tasks. In accordance to the previously results, we observed that the older the user, the higher the measured efficiency level of the accomplished task was. In other words, this means that the task was accomplished faster in respect to a predefined threshold time. However, there was an exception regarding the second task (Task#2) and for the ages between 13 and 17 years old. In that case efficiency was measured at lower levels (32%) in comparison to the previous aging class of users (37%).

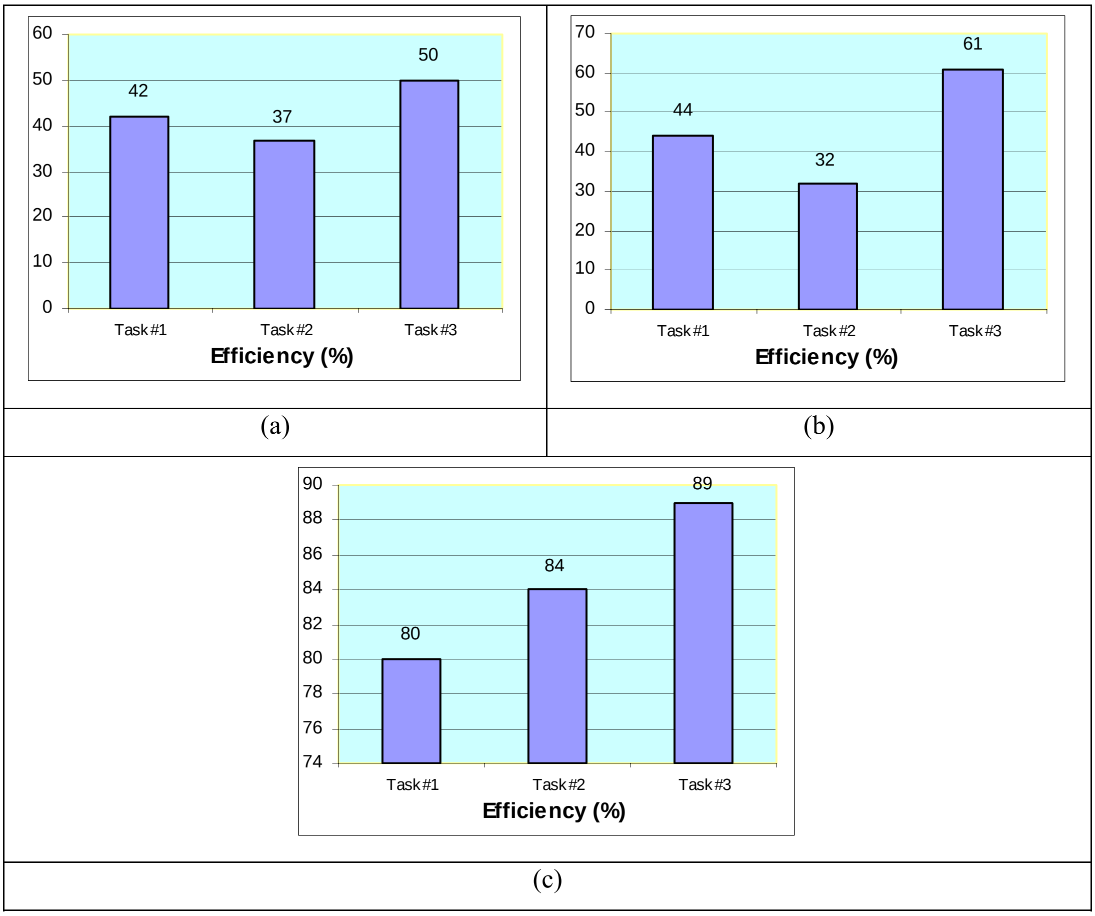
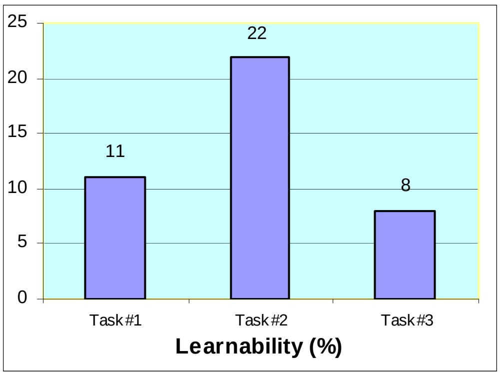
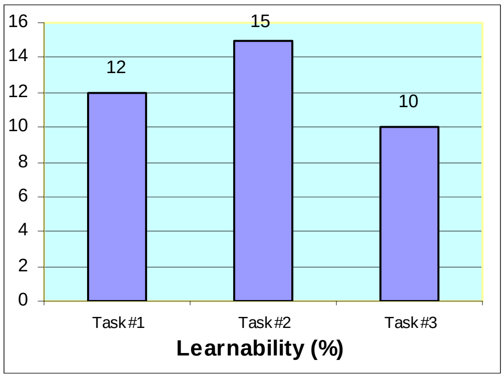


Figure 10. Efficiency measurement for users between (a) 8-12 years old, (b) 13-17 years old, (c) over 18 years old.

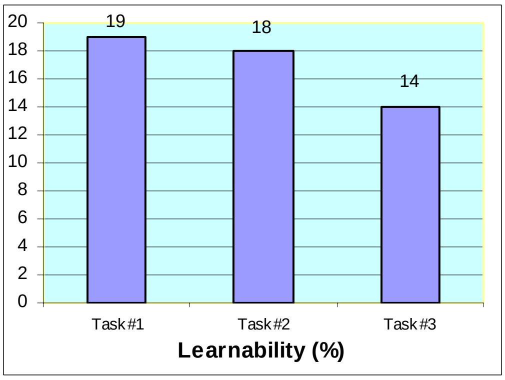
Finally, Figures 11a, 11b and 11c depict the learnability levels (in percentage) between the same user groups and for the three accomplished tasks. It was worth noticing that all aging groups presented a significant improvement in terms of the efficiency levels, when dealt with the same tasks.



(a)



(b)



(c)

Figure 11. Learnability measurement for users between (a) 8-12 years old, (b) 13-17 years old, (c) over 18 years old.

In parallel to all the above quantified metrics, the users also evaluated the proposed IPTV platform according to the following quality metrics:

- *User satisfaction:* how satisfied the users were after their interaction with the platform,
- *Simplicity:* how user-friendly the platform was,
- *Comprehensibility:* how easy the user understands messages and suggestions provided by the system,
- *Perceived usefulness:* up to what level the application accomplish its goals,
- *System adaptability:* how the systems adapts to the users' needs and preferences.

Since all these metrics were not easily quantified, the quality evaluation derived from the oral interview of all users, after the end of each evaluation session. Some positive comments and suggestions for further improvement are described in the next section.

7 Conclusions – Future Work

Towards adaptive and personalized multimedia content delivery in interactive TV environments many proposals and approaches have been discussed. This paper describes a system that takes advantage of explicit and implicit information through the users' interaction in IPTV environment. The "heart" of the system is the Adaptation Engine, which uses a personalization algorithm that weighs the relevance of the consumed content, based on the user interaction with the system.

The Adaptation Engine works not only for individual users but also reflects the profile groups of similar users and/or users with disabilities, thus decreasing the time needed for matching user patterns and profiles when forming a recommendation. This system works in conjunction with a simulation platform, acting as an interaction interface between the Internet TV architecture and the prospective viewer, providing interactivity through metadata and adaptation.

As far as the evaluation is concerned we ended up with some very positive comments, which encouraged us to continue and improve our work. Generally, the evaluation group (30 users in total) interacted quite easily with our system. Moreover, the hearing and visually impaired users evaluated as a positive point the fact that the suggested programs and multimedia content are accompanied by the necessary access and navigation services. Further positive comments derived from the oral interview were the following:

- Most users, who were quite familiar with basic use of computers and Internet navigation, rated the proposed EPG as a user-friendly application, with easy-to-use and learn functions.
- After subsequent log-ins and content reproductions, the majority of the users observed that the proposed suggestions and recommendations created a program list, which was quite relevant in respect to their preferences.
- Hearing impaired users found the functions of EPG easy to use, and their services (subscripts instructions) very helpful.

On the other hand, limitations appeared, mainly in some assistive services provided to disable users. Moreover, some hearing impaired users suggested that it would be better if the sign language, used in many provided services, was replaced by subtitles. This is because many hearing impaired users do not efficiently know the sign language. In addition, they mentioned that they felt "uncomfortable" when they consumed content with other persons who do not have hearing problems, and the assistive service was the sign language instead of subtitles. Finally, through the oral interviews we got an excellent feedback with many proposals and suggestions from the users for further improving our IPTV platform. Thus, future work includes:

- No CRID usage. CRID (Content Reference Identifier) stands as a TV-Anytime metadata allowing the user to locate and refer to content regardless of its location, whether on a particular broadcast TV channel at a specific date and time, or on a file server connected to the Internet. Although, we believed that this kind of metadata enhance our approach, it seemed that made our system more “heavy” and less user-friendly.
- The proposed TV program list should not appear automatically during the content reproduction, but only upon the user request.
- Provide different assistive information regarding the program category (e.g. news, sports, movies).
- Enhancing our approach including a Voice Server and a Text-to-Speech (TtS) engine in our architecture. The voice server will be the component responsible for the transformation of textual information to audio format and vice versa. The TtS engine will receive any text through VoiceXML files, transform it into synthetic speech and send it for reproduction in our IPTV platform.

8 Acknowledgments

This work was supported by the General Secretariat of Research and Technology (Project “Software Applications for Interactive Kids TV-MPEG-21”, project framework “Image, Sound, and Language Processing”, project number: ΕΗΓ-16). The participants are the University of the Aegean, the Hellenic Public Radio and Television (ERT) and Time Lapse Picture Hellas.

References

[1] ISO MPEG-7, part 8: Information technology -- Multimedia content description interface -- Part 8: Extraction and use of MPEG-7 descriptions, ISO/IEC TR 15938-8:2002.

[2] ISO MPEG-21, part1: Information technology — Multimedia framework (MPEG-21) — Vision, Technologies and Strategy, ISO/IEC TR 21000-1:2004.

[3] Marilly, E., Delègue, G., Martinot, O., Betgé-Brezetz, S.: Adaptation and Personalisation of Interactive Mobile TV Services, The Journal of The Communications Network, Volume **6**, Part 1, 33-39, (2007).

[4] van Barneveld, J., van SettenInvolving M.: Users in the Design of User Interfaces for TV Recommender Systems. In: UM 2003 Workshop on Personalization in Future TV, pp. 22-26, Pittsburgh, PA, (2003).

[5] Baudisch, P., Brueckner, L.: TV Scout: Lowering the Entry Barrier to Personalized TV Program Recommendations. In: Adaptive Hypermedia and Adaptive Web-Based Systems Conference, pp. 58-68, Malaga, Spain (2002).

[6] Dix, A.J., Finlay, J.E., Abowd, G.D., Baele, R.:. Human-Computer Interaction, 2nd ed., Prentice Hall Europe, London (1998).

[7] Herlocker, J.L., Konstan, J., and Riedl, J.: Explaining Collaborative Filtering Recommendations, In: ACM 2000 Conference on Computer Supported Cooperative Work, pp. 241-250, Philadelphia, USA, (2000).

[8] Setten, M. van, Veenstra, M., Nijholt, A.: Prediction Strategies: Combining Prediction Techniques to Optimize Personalization. In: UM 2002 Workshop on Personalization in Future TV, pp. 23-32, Malaga, Spain (2002).

[9] Smyth, B., Cotter, P.: A personalised TV listings service for the digital TV age. Knowledge-Based Systems, Vol. **13**, 53-59 (2000).

[10] Cotter, P, Smyth, B.: PTV: Intelligent Personalised TV Guides, In: 12th Innovative Applications of Artificial Intelligence (IAAI-2000) Conference, pp. 957-964, Austin, TX, USA, (2000).

[11] Ardissono, L., Portis, F., Torasso, P., Bellifemine, R., Chiarotto A., Difino A. Architecture of a System for the Generation of Personalized Electronic Program Guides, In: UM 2001, Workshop on Personalization in Future TV, Sonthofen, Germany, (2001).

[12] Difino, A., Negro, B., Chiarotto, A.. Telecom Italia Lab; A Multi-Agent System for a Personalized Electronic Program Guide, paper presented at Workshop on Personalizationin Future TV, Malaga, Spain (2002).

[13] Kurapati, K., Gutta, S., Schaffer, D., Martino, J., and Zimmerman, J. A Multi-Agent TV Recommender, In: UM 2001, Workshop on Personalization in Future TV, Sonthofen, Germany, (2001).

[14] Buczak, A.L., Zimmerman, J., and Kurapati, K. Personalization: Improving Ease-of-Use, Trust and Accuracy of a TV Show Recommender, In: UM 2002 Workshop on Personalization in Future TV, pp. 23-32, Malaga, Spain (2002).

[15] Dimitrova, N., Zimmerman, J., Janevski, A., Agnihotri, L., Haas, A., Bolle, R., Content Augmentation Aspects of Personalized Entertainment Experience, In: UM 2003 Workshop on Personalization in Future TV, pp. 22-26, Pittsburgh, PA, (2003).

[16] Das D., ter Horst, H.: Recommender Systems for TV. In: Workshop on Recommender Systems, Madison, WI. Menlo Park, CA (1998).

[17] Riede, C., Al-Hezmi, A., Magedanz, T.: Session and media signaling for IPTV via IMS, In: 1st International conference on MOBILe Wireless MiddleWARE, Innsbruck, Austria, (2008).

[18] Magalhaes, J., Pereira, F.: Using MPEG standards for multimedia customization, Signal Processing: Image Communication vol. **19,** 437–456, (2004).

[19] ISO/IEC 21000-7, Information technology - Multimedia framework (MPEG-21) - Part 7: Digital Item Adaptation, First Edition, 2004.
[20] Specht, D.F.: Probabilistic Neural Networks, Neural Networks, vol.**3**, no.1, 109-118, (1990).
[21] Haykin S.: Neural Networks: A Comprehensive Foundation, Macmillan College Publishing Company, New York, USA, (1994).
[22] Zhang D., Adipat B.,: Challenges, methodologies and issues in the usability testing of mobile applications, International Journal of Human Computer Interaction, 18(3), pp. 293–308, (2005).