Requirements for an Adaptive Multimedia Presentation System with Contextual Supplemental Support Media

Suzy Cutts
Software Systems Research Group
Bournemouth University
Bournemouth, UK
scutts@bournemouth.ac.uk

Philip Davies
Faculty of Technology
Bournemouth and Poole College
Bournemouth, UK
pdavies@bpc.ac.uk

David Newell
Software Systems Research Group
Bournemouth University
Bournemouth, UK
dnewell@bournemouth.ac.uk

Nick Rowe
Faculty of Technology
Bournemouth and Poole College
Bournemouth, UK
nrowe@bpc.ac.uk

Abstract—**Investigations into the requirements for a practical adaptive multimedia presentation system have led the writers to propose the use of a video segmentation process that provides contextual supplementary updates produced by users. Supplements consisting of tailored segments are dynamically inserted into previously stored material in response to questions from users. A proposal for the use of this technique is presented in the context of personalisation within a Virtual Learning Environment. During the investigation, a brief survey of advanced adaptive approaches revealed that adaptation may be enhanced by use of manually generated metadata, automated or semi-automated use of metadata by stored context dependent ontology hierarchies that describe the semantics of the learning domain.**
Keywords – e-learning, adaptive, metadata, semantic, ontology.

I. Introduction

The power of supplementing traditional materials with Video Training Modules (VTMs) in a multimedia presentation system became apparent during several years of delivering 3D modeling and animation courses. This is explained by the need for students to use processor intensive applications in these courses. Much of the content needs particular methods and techniques to be performed using complex computer applications. The methods were constrained by the design of the applications programs' interfaces so there is no alternative to showing participants specific methods, ideally visualized using VTMs. Screen activity is recorded and a recorded audio track is added where the tutor explains events on the screen. Many applications exist to create VTMs, both open source and proprietary, and there is a history of using VTMs to explain methods within the modeling and animation industry. Here, and elsewhere, their effectiveness is well established [1].

Evaluation by users provided evidence of the need to supplement VTMs with an efficient navigation system so students can re-access learning at any point in the audio/video segment. This requires provision of user controls and the structure of the presentation to be manifest to the student, for example in the form of a table of contents. Supporting text is also needed to provide greater accessibility. Student input is desirable to facilitate questions to tutors and answers from them. Furthermore, it would be advantageous to be able to select appropriate audio/video presentations from a range of pre-existing on-line sources. Ultimately such a system would operate in real time and act as a front-end for a distributed multimedia presentation system.

A brief survey of prevailing approaches to adaptive multimedia has shown that systems with personalisation requirements have begun to be developed, based for example on the Synchronised Multimedia Integration Language, SMIL. Yang and Yang [2] discuss the development of SMILAuthor, a tool that claims to offer benefits over other SMIL authoring tools. SMILAuthor generates SMIL code to spatially place objects on a presentation panel using a drag-and-drop interface. It has a visual timeline representation for the placement of events, making the generation of SMIL referring to temporal events much simpler to understand than hand coding in an SMIL document. Reducing the complexity of the process should reduce the number of errors in such code. However, SMILAuthor does not include the notion of dynamic fragmentation introduced in this paper allowing formation of multimedia material by tutors' responding to questions from students.

The structure of the paper is as follows: Section 2 gives a brief requirements specification for the proposed adaptive multimedia presentation system, Section 3 is a suggested architecture, including a brief case study, that demonstrates useful adaptation techniques, Section 4 is a short survey of future trends that multimedia presentation systems might use to store and retrieve data about multimedia content and use it to achieve meaningful user adaptations automatically. Section 5 is a conclusion and discussion of future work.

II. Requirements For Interface Design

The initial use case diagram in Figure 1 shows separate requirements for the tutor and the student. The tutor requires the minimum amount of time and effort to input necessary learning material. Initially, this is limited to sourcing and uploading the audio/video segments and being able to put them into an appropriate order. An adaptive engine within the system will extract appropriate text and timeline data from the audio/video segments and distribute this to the display panes of the interface to present the table of contents.

The student requires not only access to the audio/video segments but also a measure of control over their delivery. Being able to select and re-run segments is important for learning at the student's own pace. To enable this, an intuitive navigation system is required which sequences and orders the significant points in the presentation and displays them in a table of contents with associated support text. The ability to gain clarification on points not understood is also an essential requirement to effective learning.

A proposed prototype system shown in Figure 2 is composed of five principal parts: the main presentation panel, the table of contents panel, the supplementary text panel, the questions panel and submit button, and timeline controls for the running of the audio/video presentations.

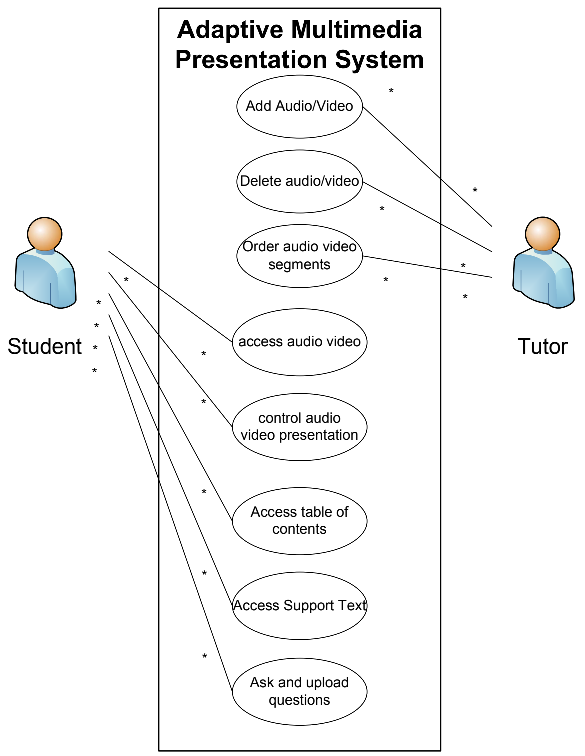

Figure 1: Use Case Diagram

A. Main Presentation Area

This contains the multimedia document which may display any combination of text, graphic, image, audio and video. It is also the primary data source from which all supplemental information is retrieved.

B. Table of Contents

The information displayed in the table of contents is automatically retrieved from the support text pane. This will

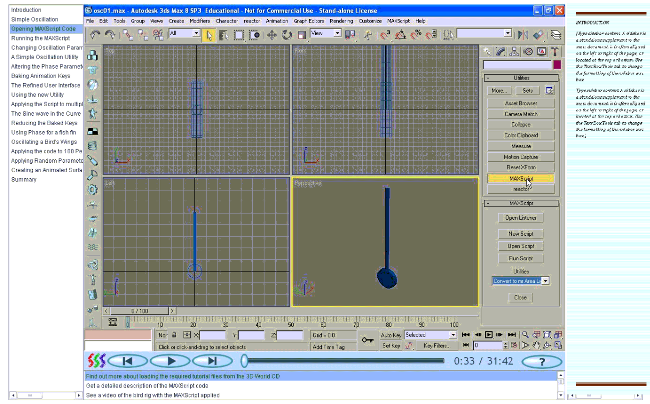

Figure 2: A proposed prototype system

require the use of intelligent knowledge storage and retrieval techniques that can structure, select and display the most useful learning material. The table of contents is presented in the form of a tree structure with a breakdown of sections. Each section title is, for example, a hyperlink to a position on the timeline, so that it is possible to jump

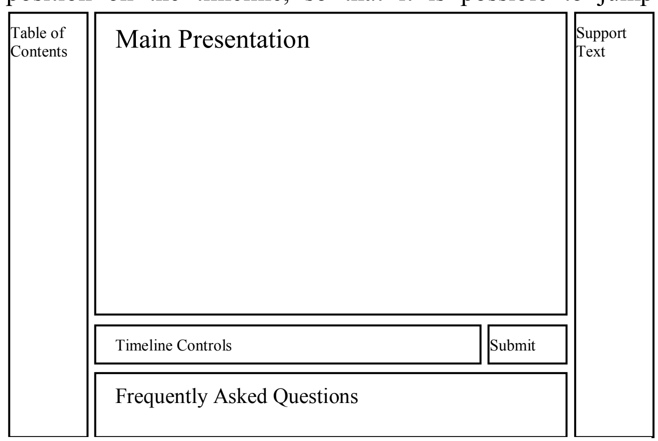

Figure 3: Schematic of the prototype system

between places within the same video/animation or sequence of them. In later developments, addition supplementary information may be provided from the main presentation area using a variety of knowledge engineering techniques including text-based retrieval, image retrieval, video retrieval, and audio retrieval to construct a more adaptable multimedia presentation. Content-based retrieval techniques vary from one element of multimedia to another, ranging from keywords for texts, colour and texture for images and spoken words for audio, for example.

C. Supporting Text

Additional supporting notes will appear in this portion of the screen. This is intended to be text that assists the user's accessibility of the learning material. It may contain hyperlinks to other timelines, i.e. will open a new window with a duplicate set of components and its own timeline. The words displayed here may be a simple transcription of the audio part of the presentation displayed in the main area which could be retrieved by voice recognition techniques.

D. FAQ and Submit Button

It was realized early that a deeper level of interaction would be required to meet the needs of a student asking a question during a class or lecture. This external interaction requires the tutor to respond to questions put by students using the system. A proposed solution is to allow the user to invoke a text dialogue with a tutor triggered by a button.

Questions are typed into the text area and submitted to the tutor with a button click. From this, an e-mail might be composed and sent. Along with the question, a unique identifier for the presentation module and a timestamp is added. The timestamp isolates the precise time in the timeline when the question was asked, allowing the tutor to track into the presentation to see the context of the question.

The user's specific question forms the basis of feedback to alert the tutor of possible clarifications in the presentation that need additional explanation. Irrelevant questions will be screened by the tutor. Once the tutor understands the question, an answer is constructed in the form of a further audio/video segment. This can be inserted into the original presentation. The text question is displayed in a FAQ region when the presentation timeline reaches the point when it was asked. The audio/video segment containing the answer can then be optionally activated by selecting the question, pausing the main presentation until the supplementary segment has been played. As more students view the modules, ask questions and gain answers, the presentation evolves by dynamically enhancing the learning resources.

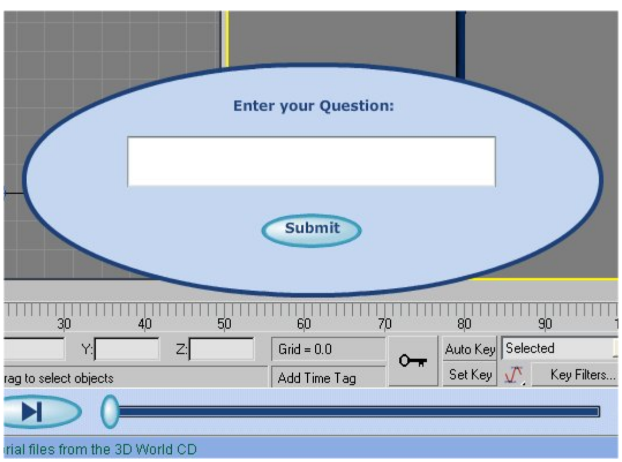

Figure 4: The submit question dialogue box

E. Media Time Line with Function Buttons

The system offers temporal interaction that allows students to move through the presentation using the time bar, offering the ability to pause a presentation using a button and clicking on the table of contents to move to a specific area. The current topic in the table of contents is highlighted in real-time so students can determine the position within the presentation. This type of interaction allows students to adjust the delivery of the presentation to suit their own learning requirements.

A graphical representation of a time line is provided, similar to a media player, representing the temporal state of the currently playing video or animation. A standard set of buttons for controlling playback will be provided. The total duration of the video/animation, or set of videos/animations which run in sequence, determines the maximum duration of the media time line.

III. Architecture

A. Media Segmentation

The tutor builds the E-learning modules by using the segmentation architecture, which provides flexible delivery. The presentation is broken down as required into multiple segments each corresponding to an individual learning object. The selection, arrangement and linking of segments will constitute the delivery of a particular VTM with a learning approach. In this way many segments could be played one after the other to view different aspects of the content. For example, screen shots within on-line learning materials may be followed by a video of a practical laboratory example.

Furthermore, in order to respond to the differing needs of learners, the linking of the media segments will involve more than just a linear arrangement. The response to student interaction requires branching capabilities within segmentation architecture [5]. Segmentation allows the selection of material according to learning need. Students may choose to view only those segments they need to see. Additionally, the system will have the ability to respond to new learners' needs not already met, or even envisioned, by currently available material. Hence the system will record and insert new media segments as required. For example, in response to a student's question for more information on a particular topic, the tutor can record a new segment.

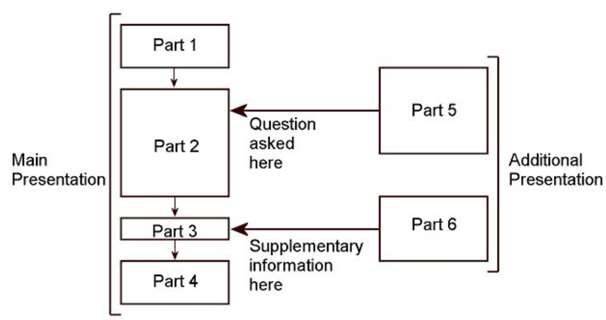

Figure 5: The timing of presentation segments

Figure 5 shows a main presentation sequence of four media segments making up a VTM. Questions asked by students at points in segment 2 and segment 3 have led to the generation of new segments 5 and 6 by the tutor which link to the main sequence at the appropriate points shown in the diagram.

MAIN SEQUENCE
<SEGMENT 1>
<SEGMENT 2>
      <SEGMENT 5>
<SEGMENT 3>
      <SEGMENT 6>
<SEGMENT 4>

This is equivalent to a multi-level list with a tree-structured architecture. Each new segment is simply added as a subsection at the appropriate place in the list which is constructed in xml. This is rendered by the system to produce a new table of contents entry and FAQ entry. When either of these is selected, a new window opens containing the video or animation explaining the answer to the query. Each term listed needs to be linked back to a point or points in the video when the term was used and is marked as a point on the timeline. Clicking the hyperlink moves the current timeline to the associated video or animation.

B. Media Player Configuration

As a single player is required to play any module, configuration is required to activate the required resources and also to give the temporal information needed to activate the table of contents entries and the FAQs. Figure 6 shows the original XML file used for configuring the system. The file has an outer main tag. The children within this are frame rate, module ID, filename, tocInfo and questions.
The filename tag contains the files to play in sequence in the main presentation area. In this case a small presentation was played before the start, ploadv2.swf. This allowed the main presentation to be preloaded while this was playing so there was no loading delay for the main presentation.

```
<?xml version="1.0" encoding="iso-8859-1"?>
<main>
   <framerate>8</framerate>
   <moduleid>V200134234</moduleid>
   <filename>
      <node name="ploadv2.swf"/>
      <node name="art02.swf"/>
   </filename>
   <tocInfo>
      <node label="Introduction" fileset="0" time="0.00" />
      <node label="Simple Oscillation" fileset="0" time="11.50" />
      <node label="Opening MAXScript Code" fileset="0" time="24.75"
/>
      <node label="Running the MAXScript" fileset="0" time="64.75"
/>
      <node label="Changing Oscillation Parameters" fileset="0"
time="109.38" />
      <node label="A Simple Oscillation Utility" fileset="0"
time="183.25" />

      ...

      <node label="Creating an Animated Surface" fileset="0"
time="1563.25" />
      <node label="Summary" fileset="0" time="1802.25" />
   </tocInfo>
   <questions>
      <node name="Find out more... " file="art01.swf" frame="88"/>
      <node name="Get a detailed... " file="art05.swf"
frame="552"/>
      <node name="See a video of..." file="art06.swf"
frame="10416"/>
      <node name="See a video of..." file="art02c.swf"
frame="12416"/>
      <node name="How can this..." file="art03.swf" frame="12560"/>
      <node name="How can the oscillation..." file="art04.swf"
frame="12640"/>
   </questions>
</main>
```

Figure 6 – The XML configuration file

A prototype design architecture satisfying these initial requirements is undergoing implementation and evaluated by the writers.

IV. ADAPTIVE AUTHORING & RETRIEVAL TOOLS

A. Development Stages

A prototype development with staged design and implementation with increasing levels of adaptation will use two Virtual Learning Environments (VLE). One VLE will be at Bournemouth and Poole College, using the well known, open source VLE named Moodle. Bournemouth University uses a localised version of the Blackboard VLE. Both VLEs have been in use for a number of years at these institutions.

The development stages will be:

1. Presentation player to display VTM content from Moodle
2. Presentation player to display VTM content from Blackboard
3. Authoring integration tool with manually entered meta data to create segmented VTMs
4. Authoring tool with automatic generation of meta data using adaptation/ontology techniques
5. Authoring tool with adaptive retrieval engine to automatically create multimedia content for presentations from generated ontology/metadata
6. Personalised adaptive multimedia presentation system based on students' assessment test results

B. Scope of User Types

The intended end user for the multimedia presentation materials in the first instance are further and higher education students. Being mainly young adults, there is no requirement to take the learning needs of children into account. However, the system needs to be intuitive since it is not intended to limit its use to students of computing or IT courses. The socio-cultural backgrounds, abilities, learning styles and accessibility issues of students attending the two institutions and initially using the system are varied. This adds to the personalisation notion to which the system is intended to adapt. By making the system accessible only through the VLE, student will understands that the system is for the facilitation of learning, not entertainment.

The second user type is the academic tutor, for whom a means of authoring presentations is being developed. The speed of material development must be comparable to other methods, for example, at least as straightforward as creating a slide presentation. This will encourage staff and students to use the system and reap the benefits. Mastering the authoring system must be quick and easy. The requirement is to create an authoring method for the presentation system that requires only a typically found user level of computer literacy.

C. Authoring Tool

The authoring tool shown in Figure 7 is crucial to the success of the presentation system. This can be evaluated by the widespread use of the system by lecturing staff and students. Success amongst staff will only occur if authoring is easy and will continue where feedback from students is widespread and positive. An authoring tool for multimedia presentations must be easy to use by non-technical teaching staff for speedy development of content [6].

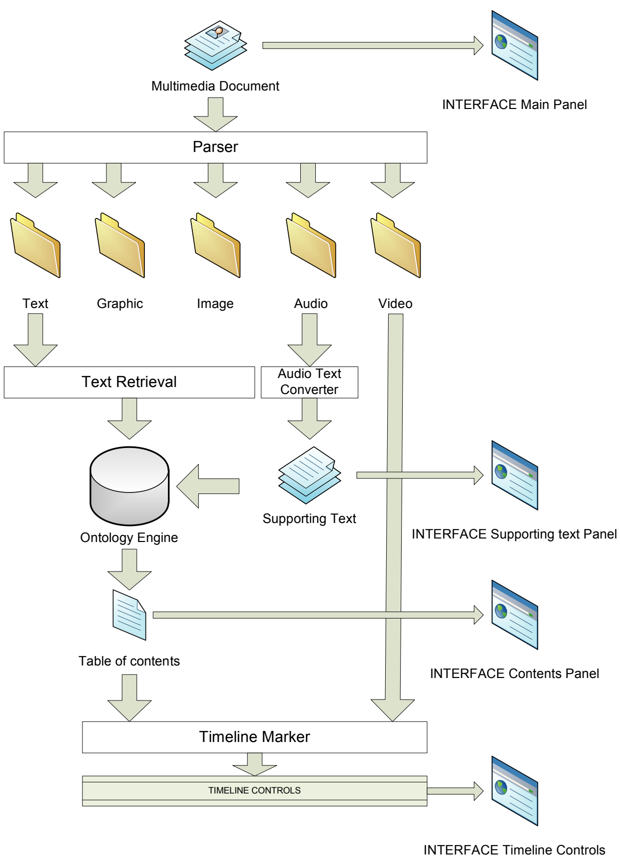

Figure 7 – Architecture of an Authoring Tool

While viewing a multimedia segment, for example, the audio of the presentation is separated and passed through a text retrieval engine which uses voice recognition principles to recover and provide text direct to the supporting text panel. Text may then be sent to the ontology engine. It uses a mixture of manual and automatically generated semantic structures that represent the conceptualisations meaningful within the context of the segment contents. The details of operation and application of ontology engines are current research areas [4]; however the required outcome is the construction of the table of contents in the form of a hierarchy of terms. In the case of a 3D visualisation tool, a heading 'rendering' might be inserted into the table of contents referring to a combination of multimedia information available in the presentation system. The timeline controls links the term 'rendering' to relevant points in the multimedia content to mark the position on the timeline. The hyperlink provides a method to access the timeline of the relevant video segment or animation.

In an effort to further reduce authoring complexity, in the simplest case, metadata describing the content of segments could be created and entered manually by a domain expert at the time of media segment creation. Future enhancements would in the limit, add capability to 'see' the frames of the video, 'see' the contents of images, 'listen' to the audio, or 'read' text. The later is the most feasible currently, for example by searching for key words in the text, building a semantic model of content known as an ontology for the problem domain, and using this to dynamically classify and construct useful content based on the meaning of available materials.

By analysing content dynamically in response to students needs in real time, the authoring tool itself would ideally be made capable of creating ontology information and using metadata. It is anticipated that the most difficult analysis would be looking for objects in videos and determining their type and meaning. However, the sports industry have analysis software for tracking the paths of moving objects such as balls on pitches and organisations involved in photography have workable face recognition systems in cameras already in use.

Further to attempting to apply knowledge engineering principles such as storage and retrieval of multimedia objects based on these knowledge engineering principles, another dimension is added when the challenge of dynamic assembly of materials based on content descriptions is extended to a fully distributed system such as the Web. Practitioners are investigating these areas actively. Henze, Dolog, & Hijdl [3] have reported on the use of a logic description language, Resource Description Formats, RDF, to guide the formation of an ontology and metadata for three types of resource – domain knowledge, user knowledge and observer knowledge. These are used for personalisation of learning in a future semantic web, although the production of quality materials in an open system is problematic.

The theoretical foundations of logic languages and frameworks such as RDF hold the promise of producing practical tools and techniques for future adaptive multimedia presentation systems but they are not fully explored yet. Providing personalised on-line learning using an ontology engine to create adaptations in a closed system, let alone an open one such as the Web, is an active and complex research area. Many writers are investigating competing methods and techniques to apply knowledge engineering based approaches to various application domains. This includes the use of multi-agent systems [7], neural networks or fuzzy logic filtering [8].

V. Conclusions and Future Work

An investigation has been undertaken into the requirements, underlying techniques and technologies needed for an adaptive multimedia presentation platform, a kind of intelligent VLE. Content selection can make use of a form of knowledge based analysis of semantic contents of multimedia segments, dynamic generation of ontology information about video segments is stored, and retrieval proceeds dynamically according to the use of the semantic data in future forms of such a system. Research issues associated with this knowledge based approach to personalisation of learning have been outlined, but are not yet fully explored. Initial findings suggest a useful architecture can be initially developed making use of an interactive segmentation method for video media. A generic framework for adapting media presentations through adding new content requested by student interaction, using a tree-branching sequencing system, rather than the usual linear sequencing system for multimedia segments, has been described. For the future, a number of promising research directions in addition to those in this work is under scrutiny. They include the design, implementation and integration of these concepts on two VLE systems for use as demonstrator tools.

References

[1] Salton, G. & McGill, M.J. (1993) *'Introduction to Modern Information Retrieval'*. McGraw-Hill.

[2] Yang, C., & Yang Y. (2003) SMILAuthor: An Authoring System for SMIL-based Multimedia Presentations. *Multimedia Tools and Applications*, 21. 243-260 Kluwer Academic Publishers, Netherlands

[3] Henze, N., Dolog, P. & Nejdl, W. (2004) *'Reasoning and Ontologies for Personalised E-Learning in the Semantic Web'*, Educational Technology & Society, 7(4), 82-97.

[4] Frensel, D., van Harmelen, F., Horrocks, I., McGuiness, D., Patel-Schneider, P. (2001) *'OIL: An Ontology Infrastructure for the Semantic Web*', IEEE Intelligent Systems.

[5] Jun Yang, Q. L. (2007). 'Retrieval of Flash™ Movies by Semantic Content: Research Issues, Generic Framework, and Future Directions. *Multimedia Tools and Applications , 31*, 1-23.

[6] Shankar Vembu, M. K. (2006). 'Towards Bridging the Semantic Gap in Multimedia Annotation and Retrieval'. *1st International Workshop on Semantic Web Annotations for Multimedia (SWAMM).*

[7] Ketter, W. Batchu, A., Berosik, G., McCready, D. (2008) *'*A Semantic Web Architecture for Advocate Agents to Determine Preferences and Facilitate Decision Making', *ACM.*

[8] Teuteberg, F.(2003) 'Intelligent Agents for Documentation Categorisation and Adaptive Filtering Using a Neural Network Approach and Fuzzy Logic' in Knowledge-based Information Retrieval and Filtering from the Web (Ed. Abramowicz, W.)*, Kluwer Academic.* 231-250