Developing a Multimedia-Based Intelligent Tutoring System

Amelia K.Y. Tong

Centre for Multimedia, School of Computing, Information Systems and Mathematics, South Bank University, London, UK

Multimedia systems lack the goal-oriented teaching approach of intelligent tutoring systems. If multimedia is integrated into intelligent tutoring systems then the combination of easy-to-access, highly-structured audio-visual information with goal-oriented intelligent tutoring increases the pedagogical effectiveness of such a system. This paper illustrates how multimedia can be incorporated into an intelligent tutoring system in the implementation of such a system which offers tutoring on zoology.

Keywords: Multimedia, Intelligent tutoring systems.

1. Introduction

Traditional Intelligent Tutoring Systems (ITSs) have mostly been developed using the knowledge-based expert system paradigm (Ellington, 1995). In an ITS, knowledge is usually organised into three models: a model to represent knowledge about the domain, a model to represent knowledge about the student, and a model to represent knowledge about how to teach. The close interconnection among these three models assumes explicit and direct links between related knowledge parts of the three knowledge representations (Angelides, 1995). There are several shortcomings with using this knowledge-based expert system paradigm alone. Firstly, knowledge based expert systems are used mainly for the purpose of representing alphanumeric data for knowledge representation in the three models, and representing the inferencing mechanism among the three models. It does not support tutorial material and the tutorial process which may require multimedia representation, such as the use of audio and video. Secondly, a knowledge-based expert system can only make hierarchical decisions based on the knowledge representation that possibly includes all available knowledge about the domain. Such ITSs cannot include non-linear links among the components. This, in turn, implies that it is difficult to include additional knowledge after the construction of the ITS is completed. Thirdly, this inflexibility of knowledge organisation imposes limitations on efficient storage and access of information.

Nowadays, computer-based training slides towards a multimedia training environment where the trainee can have access to external devices such as CD-ROM drives which can hold a wide range of random access information. However, the use of multimedia planning usually gives little thought to the expectations of any specific learning goals, less still to the activities and information essential to the achievement of the specific goal, and none at all to the formative evaluation and revision of the student's learning against a specific goal (Angelides and Dustdar, 1997).

The purpose of this paper is to illustrate how traditional ITSs can be integrated with multimedia, to provide a multimedia intelligent tutoring environment which combines structure, control, knowledge representation, inferencing and problem solving capabilities, together with a wider means of delivering tutorial material, and more flexibility in non-linear access to information. Neither traditional ITSs nor multimedia alone is sufficient to offer such an environment efficiently. ITSs with multimedia support may be coupled with the features which the multimedia component may contribute to the overall traditional ITS environment: an enhancement to the system-student interaction through the active use of audio and video, locating and retrieving information either in the knowledge base or the multimedia base, alternative methods of presenting the tutorial material to the student, an annotation facility allowing the background knowledge and explanations to be captured from the expert that would otherwise be entered in the structured knowledge base of the ITS or that would not fit in the computational framework of the ITS.

The paper is organised as follows: first it explains how multimedia is represented in ITSs. The paper then proceeds to describe the development of ARISTOTLE, in terms of the structure of its domain model, its student model and its tutor model, to demonstrate how multimedia is incorporated into the standard architecture of an ITS.

2. Intelligent Tutoring System Representation of Multimedia Material

The main characteristic of a multimedia-based ITS (MITS) is its ability to deal not only with alphanumeric data, but also with still and full-motion video, audio, graphics and animation (Furht and Milenkovic, 1995; Rodriguez and Rowe, 1995; Steinmetz and Nahrstedt, 1995). MITSs therefore have to handle new data types and their relationships together with the traditional ones, i.e. retrieving and processing mechanisms for static media, such as graphics and text, as well as for dynamic, time-variant media, such as video and audio.

2.1. Syntax and Semantics in Video and Audio

MITSs should achieve this by representing all information uniformly, as a bit stream. However, much work has been dominated by this bit stream: video has been developed so far as a technology of images, and audio as a technology of signals (Liou, 1991; Chang and Hsu, 1992; Steinmetz and Nahrstedt, 1995; Meyer-Boudnik and Effelsberg, 1995). These issues of syntax alone cannot fully address the issue of how to use video and audio effectively within a MITS environment. This is because video and audio will only become effective parts of everyday computing environments when they can be used with the same ease as text. We do not use video just because the images are steady, or audio because it sounds stereophonic. We use these media for their content. This is a semantic issue. Without knowledge of its content a bit stream will remain uninterpreted. To use and interact with it, the bit stream has to be converted into a form that can be understood. It is therefore important to distinguish between the syntax and semantics of video and audio.

Distinguishing between syntax and semantics in MITSs enables a distinction to be made between pixel and semantic representations in video, and signal and semantic representations in audio. Multimedia syntax includes text representation through ASCII code, video represented through formats such as MPEG, and audio represented through wave and other formats. In contrast, multimedia semantics is concerned with the meaning of the content of the video and audio (Agius, 1997; Agius, 1998).

The discussion so far suggests that an MITS requires the following:

- Comprehensive information about the video and audio to be used so that the system is able to use various segments of the video and audio according to their content.
- Structured information about the video and audio, suitable for use by pedagogical components of the ITS.
- Links between the segments-related content, so that they may be used collectively within the ITS when necessary.
- Tight links between video and audio, the content-related information, and the knowledge held within the system to be taught to the student, so that all the three can be used conjointly.

The next section illustrates how the above requirements can be met by using the multimedia frame structure to represent multimedia in MITs.

2.2. Multimedia Frames

Agius (1997) proposes the use of multimedia frames (m-frames) to represent video and audio in order to satisfy what an ITS requires to represent multimedia. M-frames provide an object-oriented manner by which the syntactic and semantic information can be used together. Since both syntactic and semantic information

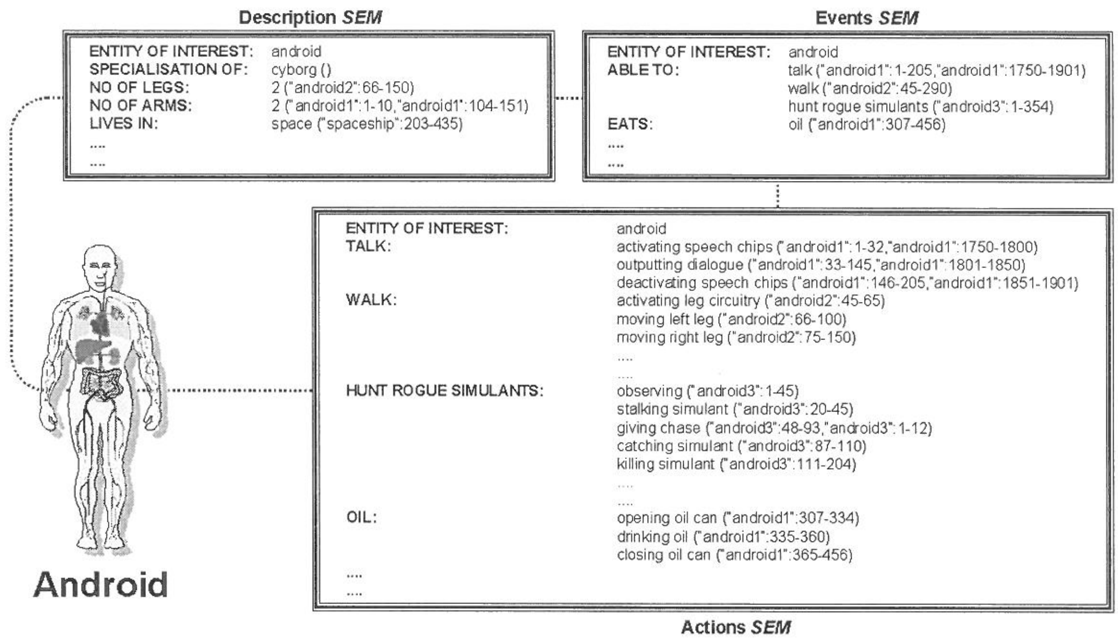

Fig. 1. A sample SEM.

are required, two types of m-frames are distinguished: syntactic m-frames (SYMs) which model the syntax of each frame's content, and semantic m-frames (SEMs) which model the semantics of several shots content. A SYM models the content of a video/audio frame. Each video and audio within an ITS therefore has a group of syntactic m-frames associated with it. The content modelled by SYM are the objects present, together with their on-screen coordinates, and the spatial relationships between the objects. Since audio does not have the meaning on an individual frame-by-frame basis but has the meaning over time, the information concerned with audio meaning in time is modelled by SEMs.

A SEM provides information about the semantic content of various segments of video and audio frames that are related to a particular concept pertinent to the ITS. It is therefore within the SEMs that shots are defined. A SEM also provides semantic information that is not related to a media segment within the ITS. In this way, all information related to an entity of interest to the ITS, i.e. both content-based and non-content based information, are kept together. An entity of interest to the ITS is an object in the domain to be taught. Each entity of interest is represented by a collection of three SEMs: the Description SEM describes the entity of interest, the Events SEM models the events that are associated with the entity of interest, and the Actions SEM models the constituent actions of the events modelled in the Events SEM. The SEMs therefore group together the media segments that are related to an entity of interest. Figure 1 shows a conceptual representation of Description, Events and Actions SEMs for an Android entity of interest, which is a specialisation of a Cyborg entity of interest.

Each slot within the SEM represents a particular perspective on the multimedia content, with the slot values representing more specific instances. The slots and values of the SEMs are defined by the domain and by the entities of interest, unlike SYMs, which have a predefined format. Shots are defined for instances within the SEMs, with multiple shots separated by commas within the figure.

This section has described how video and audio can be represented in an ITS which includes multimedia material. The next section describes the development of ARISTOTLE, a MITS, in terms of how the multimedia material is prepared, and how it is incorporated in ARISTOTLE's architecture, in terms of the domain model, the student model and the tutor model.

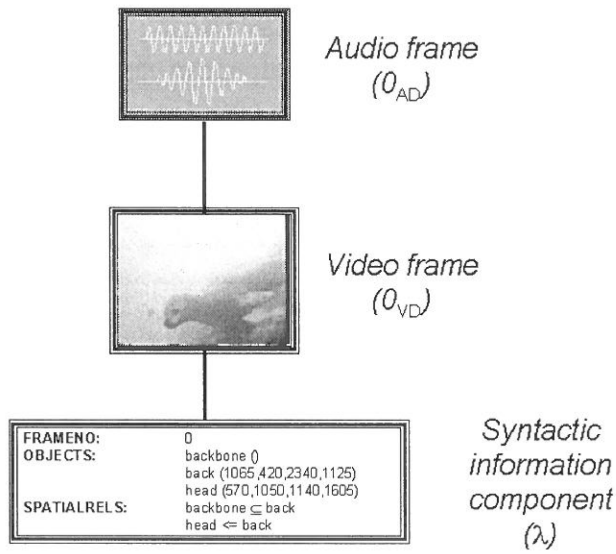

Fig. 2. A SYM in ARISTOTLE.

3. ARISTOTLE

ARISTOTLE was developed with Asymetrix Multimedia Toolbook 4.0 under Microsoft Windows. The system tutors the knowledge of basic zoology to young school children. ARISTOTLE has been designed to fit within the framework of the National Science Curriculum for England and Wales (Department of Education, 1995). The Curriculum encourages implementation and use of multimedia because it is useful for teaching about the visual and aural phenomena, such as movement, observable differences between living things, growth, finding different animals in different habitats, and distinguishing variation in the noises of different animals.

ARISTOTLE's domain is that of zoology, therefore the entities of interest to the system are animals. Animals were chosen and grouped into classes of vertebrates and invertebrates. Vertebrates and invertebrates were further divided into mammals, reptiles, arthropods and so on. Then it was decided which animal events and actions would be used in teaching. The events for the chosen entities centred on what the animal or animal class was able to do, for example, hunting, while the action broke down the events into their constituent activities, for example, catching prey. Suitable video and audio segments were then captured and edited into clips. All the clips were added to a single multimedia base using the Clip Manager, front-end software tool developed by Agius (1997) in Asymetrix Multimedia Toolbook that adds, modifies and deletes clips within a given multimedia base.

Once the video and audio footage has been collected, spatial relationships of the objects within the footage is modelled. This information is then used to implement the SYMs. SYMs were implemented using the SYMulator front-end software tool (Agius, 1997) in Asymetrix Multimedia Toolbook. Figure 2 provides a conceptual representation of a SYM from ARISTOTLE.

The SEMs in ARISTOTLE were then constructed according to the description, actions and events about each entity. The SEMulator front-end software tool (Agius, 1997) in Asymetrix Multimedia Toolbook was used. First the perspective and instances were created for each of the Description, Events and Actions SEMs. Then appropriate shots were defined for the various instances with these SEMs.

After the multimedia material has been prepared, ARISTOTLE is now ready to incorporate them into its otherwise traditional ITS architecture. Figure 3 illustrates ARISTOTLE's architecture. Multimedia is incorporated into all of the three primary components of traditional ITSs.

All of ARISTOTLE's SEMs, with the exception of the remedial strategy and teaching strategy SEMs, are stored as Borland Paradox databases

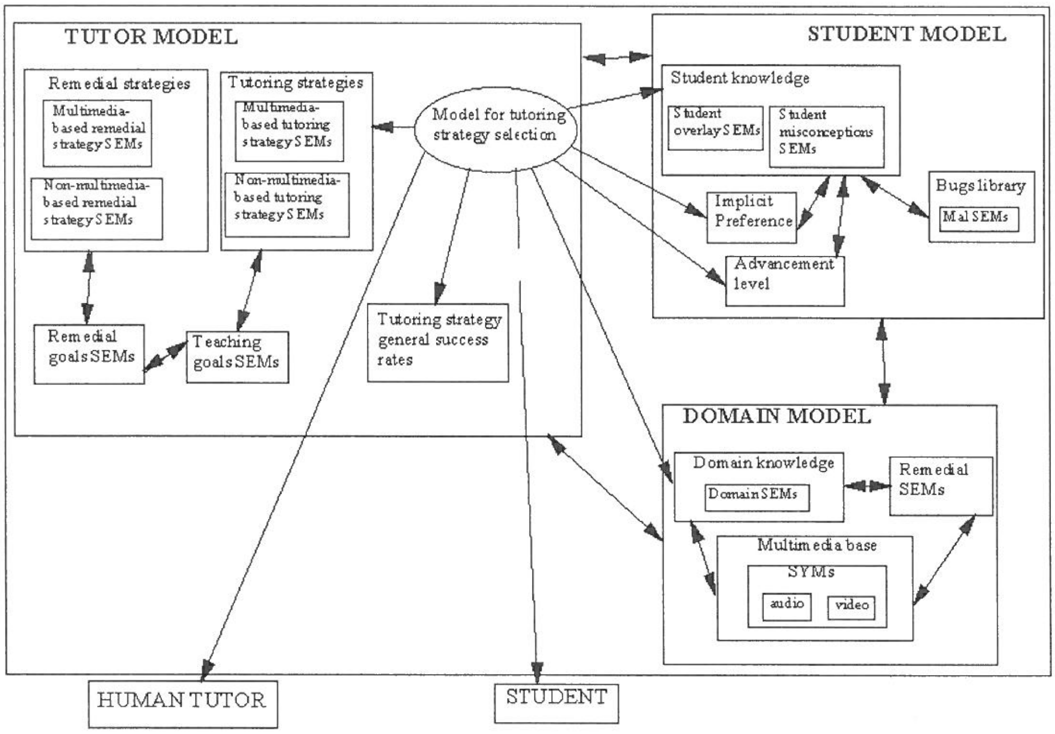

Fig. 3. ARISTOTLE's architecture.

and each consists of two fields: the perspective, which is of type *alphanumeric*, and the instances, which are of type *memo*. The remedial strategy and teaching strategy SEMs are procedural in nature and therefore are stored as an Openscript code in the Multimedia Toolbook.

3.1. The Domain Model

The domain model of an ITS should contain the domain knowledge which represents the problem solving expertise (Van Jooligen, 1995), i.e. the specific knowledge of the domain to be taught. This knowledge can be applied to solve problems in the specific domain when combined with a rule-based inference processor. The nature of the knowledge stored in the domain model determines the content of a tutorial discourse, the goal structure that may influence the system's selection of examples, questions and states made at different points of the tutorial discourse. In a traditional ITS, the domain model is represented by the domain knowledge and may be remedial knowledge in addition. In the case of ARISTOTLE, the domain model also includes a multimedia component. The multimedia base consists of the SYMs and their associated raw video and audio.

While traditional ITS's domain knowledge may only contain plain text, ARISTOTLE's domain knowledge consists of the SEMs which contain multimedia information. Figure 4 provides a conceptual representation of the Cheetah domain SEMs.

The links to the SYMs that are associated with the shots used within the cheetah domain SEMs are dynamic and are established during the course of interaction. These links are established by linking the SYM database for the shot currently being delivered with the entity of interest at hand through the manipulation of a database alias. The Borland Paradox Engine uses an alias to refer to a database table, as a substitute for the table's physical name. Thus, when a new SYM database is required, the link to it is established by changing the physical name associated with the SYM database alias, while leaving the alias name itself intact.

Links to the domain SEMs that are one level up in the hierarchy, for example, the mammal domain SEMs in the case of the cheetah represented in Figure 4, are also established dynamically. Such links are set up by the creation of three database aliases, which serve as links to

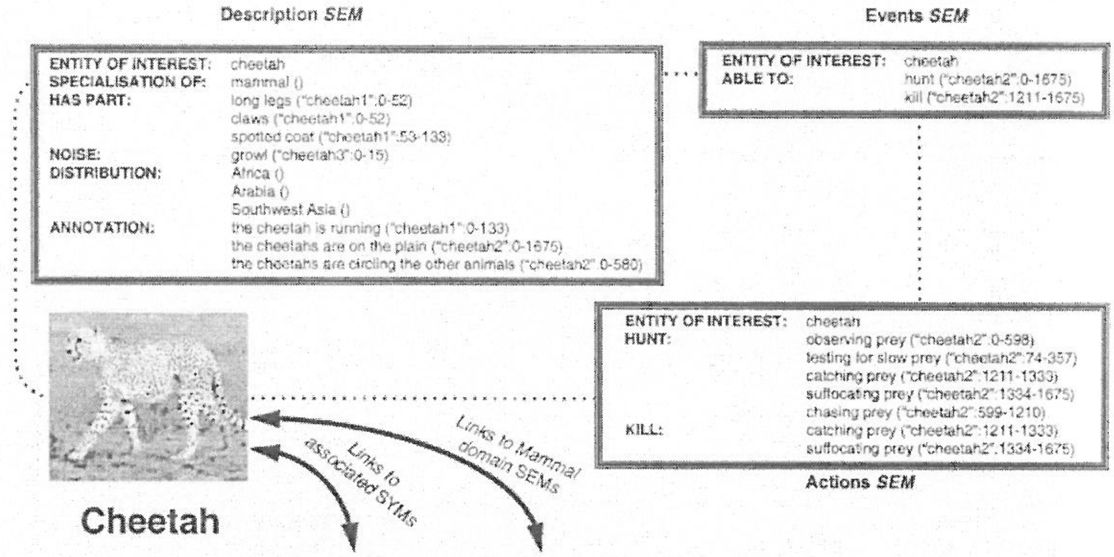

Fig. 4. ARISTOTLE's cheetah domain SEM.

the Description, Events and Actions SEMs (i.e. tables) for the required entity of interest.

The remedial knowledge stores the information used by the tutor model to remedy the student when he/she does not provide the correct answer to a question and when assistance to the student is necessary. ARISTOTLE's remedial knowledge is a hierarchy of remedial SEMs, which mirrors the domain SEMs hierarchy. The remedial SEMs are oriented towards the provision of remediation for the student. The remedial SEMs therefore model the information intended to guide the student towards remedial goals, thus the teaching goals, rather than to the reflecting perspectives on the multimedia content, as the domain SEMs do.

Procedures in the domain model, as in the other models in ARISTOTLE, do not only manipulate the information related to text, but also the co-ordinates in video clips, for example. These procedures are concerned with various tasks such as determining which objects within a given SYM match a given set of co-ordinates. This is used to determine what object the student has clicked on. Procedures within the domain model also provide for the student and tutor models.

3.2. The Tutor Model

The tutor model in an ITS gives form and meaning to the system. The characteristics of the tutor model (McManus and Aiken, 1995) include: control over the tutorial discourse, i.e. the selection and sequence of the tutorial material to be presented to the student; control over instruction, i.e. the process of the actual presentation of the tutorial material; capabilities for responding to a student's questions about instructional goal and content; and measures for determining when a student needs help and for delivering the appropriate help. The aim of this model is to circumscribe the nature of tutoring and to implement tutoring as a solution to the educational communication problem. As an ITS, ARISTOTLE's tutor model contains knowledge on teaching goals, tutoring strategies, remedial goals, remedial strategies, tutoring strategy general success rates, and a model for tutoring strategy selection. As an MITS, multimedia information and manipulation are reflected in all parts of knowledge in the ARISTOTLE's tutor model.

Teaching goals determine what the student is to be taught. There is one teaching goals SEM for each entity of interest in the domain. Figure 5 shows the cheetah teaching goals SEM.

Each goal within the SEM consists of the name of the perspective that the goal is concerned with, the minimum number of instances that the student has to name in order to satisfy the goal, and the tutoring strategies which may be used in attempt to reach the goal. The order in

ENTITY OF INTEREST:	*cheetah*
GOAL 1:	*SPECIALISATION OF 1 (1,2,3)*
GOAL 2:	*HAS PART 2 (M1,M2,M3)*
GOAL 3:	*NOISE (M1,M2,M3)*
GOAL 4:	*ABLE TO 1 (M1,M2,M3)*
GOAL 5:	*DISTRIBUTION 1(1,2,3)*
↓ Links to associated tutoring Strategies	↓ Links to associated remedial goals SEM

Fig. 5. ARISTOTLE's cheetah teaching goals SEM.

which the tutoring strategies are to be used is determined by the model for tutoring strategy selection. Multimedia is included here through the link associated with the tutoring strategies. Both multimedia-based and non- multimedia-based tutoring strategies are used in ARISTOTLE. The links to the associated tutoring strategies are dynamic and are established during the course of interaction by calling the appropriate tutoring strategy SEM. The links to the associated remedial goals SEM are established when the tutor model first begins to teach about the entity of interest associated with the teaching goals SEM. The link is established through the use of a database alias to the remedial goals SEM.

Tutoring strategies in ARISTOTLE refer to any mode or modes of action or specific activity that facilitates the accomplishment of a tutorial goal. To accommodate multimedia in the system, ARISTOTLE uses three non-multimedia based and three multimedia based tutoring strategies. Non-multimedia based tutoring strategies are used to tutor information about an animal or an animal class that cannot be taught through the use of multimedia. For example, animal categorisations are taught with the text, because no suitable audio or video shots exist within the ARISTOTLE's domain model. Multimedia-based tutoring strategies are those that are concerned with the use of multimedia within the tutorial discourse. These tutoring strategies provide different ways to tutor a particular subset of knowledge with the appropriate use of multimedia. Most of the ARISTOTLE's tutorial discourse uses multimedia-based tutoring strategies. For example, when tutoring body parts or activities of a particular animal, appropriate shots of that animal will be used.

Tutoring strategies are represented in tutoring strategy SEMs in ARISTOTLE. A tutoring strategy SEM in ARISTOTLE consists of five perspectives. They are: type, teaching tactics, check teaching tactics, testing tactics and operations. Instances of the teaching tactics perspective enable ARISTOTLE to teach according to the teaching goals. Instances of check teaching tactics perspective enable ARISTOTLE to check the previous teaching for missing concepts and misconceptions. Instances of the testing tactics perspective enable ARISTOTLE to test the student according to the teaching goals. Instances of the operations perspective enable ARISTOTLE to determine how well the student has provided an answer to a question, and to record the answer in the student model. Figure 6 shows a representation of a multimedia-based tutoring strategy SEM in ARISTOTLE.

Remedial goals are used to provide remedial assistance to the student. The amount of assistance given is governed by the model for tutoring strategy selection. There is one remedial goals SEM for each teaching goals SEM in ARISTOTLE. Each remedial goal in a remedial goals SEM consists of a number of subgoals, each of which has an associated remedial strategy, which is used to carry out the remediation. Links to associated remedial strategies and associated teaching goals SEM are dynamic and are established during the course of interaction, by calling the appropriate remedial strategy SEM based on the remedial strategy called for. ARISTOTLE has three non-multimedia-based and three multimedia-based remedial strategies. Remedial strategy SEMs have a type perspective which describes the strategy, and a tactic perspective which enables ARISTOTLE to present remedial information to the student.

```
TYPE:                        Multimedia-based Question/Answering
TEACHING TACTICS:            sCurrentInstanceNo of this book = 0
                             send doNextMultimediaInstance
CHECK TEACHING TACTICS:      send askQuestion_Multimedia_1 FALSE
TESTING TACTICS:             send askQuestion_Multimedia_1 TRUE
OPERATIONS:                    if pClicking = false then
                                 send NonMultimediaBasedTeachingStrategy_1_Operations
                               else
                                 sysCursor = 4
                                 sNoOfRightAnswers of this book = 0
                                 If ASYM_ItemOffset(sCurrentPerspective of this book,sEvents
                                 of thisbook)<> 0 OR \
                                 ASYM_ItemOffset(sCurrentPerspective of this
                                 Book,sDescriptions of this book) <> 0 then
                                   clear sEventsNamed of this book
                                 end if
                             ...
                             ...
```

↓

Links to associated teaching goals SEMs

Fig. 6. A multimedia-based tutoring strategy SEM in ARISTOTLE.

General success rates of the tutoring strategy are calculated from the ARISTOTLE's student model and recorded in the tutor model as one of the pieces of information the model for tutoring strategy selection refers to when selecting the next tutoring strategy to use. Each tutoring strategy, either multimedia-based or non-multimedia based, has attached to it one weight showing how successful it has been in teaching in general, and one weight for testing. The model for tutoring strategy selection in the tutor model retrieves information from within the tutor model, the student model and the domain model, in order to decide which tutoring strategy is to be used at any given moment of time. The model for tutoring strategy selection can cope with both multimedia and non-multimedia information.

3.3. The Student Model

The student model in an ITS constitutes a repository for information about a system-using student. This includes the knowledge, concept and skills which the student has obtained, those for which the student has shown some understanding, as well as misconceptions acquired. The student model consists of the student's history of responses and problem solving behaviour. This information is gathered from the system-student interactions throughout the tutorial discourse. From such information, cognitive needs and the most suitable tutoring strategy for the student may be deduced by the model for tutoring strategy selection.

ARISTOTLE's student model represents the student knowledge, the student's implicit tutoring strategy preference and the student's advancement level. All of which have to handle representation involving multimedia, except the advancement level. Student knowledge in ARISTOTLE is modelled in the form of the student overlay knowledge, student misconceptions and complemented by a bugs library.

The student overlay knowledge is represented by the student overlay SEMs. The student overlay SEMs mirror the structure of the domain SEMs, but include different information. Each instance in a student overlay SEM records the correct answer the student has provided, the strength of the knowledge acquired during the teaching mode and during the testing mode, the successful tutoring strategy used to elicit this response, and the shots that were used if a multimedia-based tutoring strategy was used. Student misconceptions are represented with the student misconception SEMs. These are similar to the student overlay SEMs, but the appended numbers indicate the seriousness of the bad knowledge, as opposed to the strength of the acquired knowledge. The bugs library contains student's common misconceptions about the domain. The bugs library is represented with mal SEMs with common mistakes about the class or values for particular animals. The class misconceptions represent cases in which a student may feel that certain animals look or sound alike. The student may therefore wrongly perceive one animal as belonging to a generic animal class of which it is in fact not a member.

The implicit tutoring strategy preference is deduced from the student knowledge and recorded in the student model. The model for tutoring strategy selection in the tutor model analyses the student knowledge to infer which tutoring strategy is suitable for the particular student at a given time, demonstrated implicitly by individual success rates of the different tutoring strategies applied with the student. The advancement level in the student model represents two stages of the student cognitive progress: novice and advanced. The advancement level represents the level of the student and also how far along the particular expertise level he/she has got. The model for tutoring strategy selection in the tutor model takes the advancement level into account when choosing among the tutoring strategies. This is the component in the student model which is not directly affected by the fact whether the ITS includes multimedia or not.

4. Conclusion

This paper has demonstrated, by presenting ARISTOTLE, how a traditional ITS can include multimedia in its tutorial discourse to provide a more flexible way for student-system interaction, storing and manipulating information. Such systems require detailed information about the content of the video and audio to be used for teaching, so that these multimedia components can be actively used in the tutorial. For this purpose m-frames are used to integrate them into the runtime process of the system. M-frames integrate content-related information about the video and audio segments pertinent to the pedagogy of MITSs. The video, audio and m-frames are therefore conceptually joined together and are used conjointly. M-frames fully represent a variety of perspectives on the multimedia data, including intangible events as well as objects. Contents within the many video and audio are interrelated through the use of the m-frame hierarchy (Agius, 1997; Agius, 1998).

ARISTOTLE itself, with or without multimedia, is an ITS employing the traditional ITS architecture that includes three primary models representing the domain, the student and the tutor. In addition, it includes a model for tutoring strategy selection to determine which tutoring strategy is to be used for a particular student, which in turn determines how the tutorial is to be delivered. The model for tutoring strategy selection is derived by uncovering the underlying principles of tutoring strategy selection suggested or used among the current practices in machine and human teaching.

References

H.W. AGIUS, A full-scale semantic content-based model for interactive multimedia information systems. PhD Thesis, London School of Economics, London, 1997

H.W. AGIUS, A survey of semantic content-based multimedia models. *Journal of Computing and Information Technology*, this issue (1998).

M.C. ANGELIDES, Developing hybrid intelligent tutoring and hypertext systems. *The New Review of Hypermedia and Multimedia*, **1** (1995), 67-106.

M.C. ANGELIDES, S. DUSTAR, *Multimedia information systems*. Kluwer Academic Publishers, Boston, 1997.

S.K. CHANG, A. HSU, Image information systems: Where do we go from here? *IEEE Transactions on Knowledge and Data Engineering*, **4**(5) (1992), 431-442.

DEPARTMENT OF EDUCATION, *Science in the national curriculum*. HMSO, London,1995.

H. ELLINGTON, Educational innovation — where are we now? A critical review of recent developments. *Innovation and Learning in Education*, **1**(1) (1995), 15-20.

B. FURHT, M. MILENKOVIC, Introduction. *In A guided tour of multimedia systems and applications* (B. Furht, M. Milenkovic, Eds.), (1995) pp. 1-3. IEEE Computer Society Press, Los Alamitos, CA.

T. KHAN, Y.J. YIP, Pedagogical principles of case-based CAL. *Journal of Computer Assisted Learning*, **12** (1996), 172-192.

M. LIOU, Overview of the pX64 kbit/s video coding standard. *Communications of the ACM,* **34**(4) (1991), 59-63.

M.M. MCMANUS, R.M. AIKEN, Monitoring computer-based collaborative problem solving. *Journal of Artificial Intelligence in Education*, **6**(4) (1995), 307-336.

T. MEYER–BIOUDNIK, W. EFFELSBERG, MHEG explained. *IEEE Multimedia*, **2**(1) (1995), 26-38.

A.A. RODRIGUEZ, L.A. ROWE, Multimedia systems and applications. *Computer*, **28**(5) (1995), 20-22.

R. STEINMETZ, K. NAHRSTEDT, *Multimedia: Computing, communications and applications.* Prentice Hall, New Jersey, 1995.

W.R. VAN JOOLIGEN, QmaPs: Qualitative reasoning for simulation learning environments. *Journal of Artificial Intelligence in Education*, **6**(1) (1995), 67–89.

Received: January, 1998
Revised: March, 1998
Accepted: May, 1998

Contact address:
Amelia K.Y. Tong
Centre for Multimedia
School of Computing
Information Systems and Mathematics
South Bank University
103 Borough Road
London SE1 0AA
phone: +44 (0)171-815 7462
e-mail: tonga@sbu.ac.uk
UK

AMELIA K.Y. TONG is a lecturer in the School of Computing, Information Systems and Mathematics at South Bank University. She holds a BSc degree in Computing, an MSc and a PhD in Information Systems, all from the London School of Economics. Her research areas are intelligent tutoring systems and multimedia.