Theory Based Design & Evaluation of Multimedia Presentation Interfaces

Peter Faraday

Submitted for part of examination for Doctor of Philosophy

Centre for HCI Design,
School of Informatics,
City University,
London.

September 1998

Declaration

The Author grants powers of discretion to the University Library to allow this thesis to be copied in whole or in part without further reference to the Author.

Abstract

Multimedia (MM) Applications currently suffer from an ad hoc development process. This places the usability and effectiveness of many MM products in doubt. This thesis develops a theoretically motivated design method and tools to address these problems.

The thesis is based on an analysis of the cognitive processes of attending to and comprehending an MM presentation. A design method is then developed based on these cognitive processes. The method addresses the problem of selecting media to present information requirements, how to design the media to effectively deliver the desired content, how to combine verbal and visual media successfully, and how to direct the user's attention to particular part of the presentation.

A number of studies are then presented which provide validation for the method's claims. These include eye tracking to analyse the user's reading / viewing sequence, and tests of expert and novice recall of MM and conventional text / speech presentations. A set of re-authoring studies show that application of guidelines improves retention of the content.

The method is supported by a design advisor authoring tool. The tool applies the guidelines using a combination of a critiquer and expert system. The tool demonstrates that the guidelines are tractable for implementation, and provides a novel approach to providing authoring advice.

Both the method and the tool are also validated in case studies with novice users. These demonstrate that the method and tool are both usable and effective.

Chapter 1

Introduction

'Multimedia' presentation describes the use of text, speech, image and animation to deliver information, mediated by computer. One of the areas in which multimedia presentations are likely to be important is in the delivery of task based expository ('how-to-do-it') information. [DataMonitor 95] predicts that the worldwide professional market for multimedia training and education will grow from $186.6 million in 1995 to $7.7 billion in 2005, with the home education market growing from $124.9m to $6 billion in the same period.

However, in order to be of real use, multimedia presentations need to be shown to be effective in delivering the required information. There are several problems which make this question important and of research interest.

First, little work has directly addressed the effectiveness of multimedia products. If multimedia presentations are to be adopted, they must be shown to be an improvement against existing presentations e.g. in expository materials, the user should gain a better understanding of the content and thus have a better subsequent recall. It is also important that the reasons why multimedia presentations are effective are explored. Studies are thus required which contrast existing text presentations with multimedia; and which also contrast different designs of multimedia presentation.

Second, there are few existing methods to guide the design of multimedia products. Design methods allow the process of constructing a product to be reviewed and critiqued. It would be of little use to show that multimedia presentations can be effective, but not to rationalise the decisions which had been taken in producing the presentation. A method is required which clearly states how a particular design was produced.

Third, iterative design of multimedia is expensive. Because of the costs of creating content, it is inefficient to repeatedly produce, test and re-author multimedia presentations. Design methods are one solution to this problem, providing best practice and heuristics which prevent common design mistakes. A method should make the finished product less haphazard and more likely to be effective. Without a systematic approach, diagnosing problems becomes fortuitous; furthermore, changing one part of the presentation may solve certain problems, but could equally lead to others.

Finally the design of effective multimedia is not just important as an industrial problem, and also as an academic concern. Designing multimedia presentations concerns issues in human cognition, such as attention, working memory and comprehension. By studying how combinations of media are processed, some of these complex issues may be further explored to make design more usable.

In summary, the goal of this thesis is thus to place multimedia presentation design within a stronger theoretic grounding, and show how such knowledge can be used to improve the effectiveness of presentations for delivering their content.

The remainder of this chapter will review existing studies which have addressed how to design multimedia presentations. The implications of the studies are analysed, and any shortcomings noted. The chapter ends with a summary of how this thesis attempts to improve multimedia presentation design.

1.1 Existing approaches

A number of existing areas were initially reviewed as sources of possible solutions to these problems : interface design, instructional design and educational psychology, and studies of multimedia design and automated presentation design tools.

1.1.1 Interface design

One potential source of design guidelines for multimedia is interface design. An example of the type of general HCI guidelines available are those provided by [Smith & Mosier 84]. These offers guidelines for design of user interface software in six functional areas: data entry, data display, sequence control, user guidance, data transmission, and data protection. The following is a sample from the 162 display guidelines provided :

i) At any step in a transaction, ensure that whatever data a user needs will be available from the display
ii) Display data to users in directly usable form
iii) Use short simple sentences
iv) Begin every display with a title and header, describing briefly the content or purpose
v) Consider colour coding for applications where the user must rapidly distinguish among several categories of data
vi) For size coding, a large symbol should be 1.5 times the size of the next smallest symbol

The main problem with the [Smith & Mosier 84] interface guidelines is their lack of direct applicability to multimedia design. They fail to deal in any detail with the choice of media, its design, or the user's cognitive resources required to process it.

Whilst the [Smith & Mosier 84] approach favours a large number of guidelines, a simplified method is offered by [Nielsen 92], termed 'heuristic evaluation'. This is designed to be used early in the design cycle to find usability problems. It is composed of ten heuristics, which draw together general principles for design. Those most relevant to multimedia design are given below.

i) Use simple and natural dialogue : 'interfaces should match the user's task in as natural a way as possible'; 'The ideal is to present exactly the information the user needs - and no more- at exactly the moment needs it';'Principles of graphics design can also help users prioritise their attention to a screen by making the most important elements stand out.'

ii) Speak the users language : 'aim at good mappings between the computer display of information and the users' conceptual model of the information'

iii) Minimise user memory load : 'displaying too many objects and attributes will result in a loss of salience for the ones of interest to the user, so care should be taken to match object visibility as much as possible with the user's need.'

iv) Be consistent : 'The same information should be presented on the screen in the same location on all screens and dialogues'

As with [Smith & Mosier 84], these heuristics only offer limited specific advice for designing multimedia interfaces. They do draw out the need for media design in graphics, but fail to offer any concrete advice as to how this is to be achieved.

Perhaps the most succinct summary of the available guidelines for multimedia interfaces within HCI is provided by [Markus 92]. In reviewing the impact of design guidelines upon 'spatial displays' involving the use of pictures and animation, [Markus 92] provides the following advice : 'Developers and users will need to make sophisticated decisions about the use of typography, symbols and pictograms, colour, layout and illustration, animation effects, and sequencing of information to produce optimum spatial effects.' It is concluded that 'In a few cases libraries of shapes, organising structures, or colour templates for a limited variety of applications may exist, but in most cases the quality of the final design will depend upon the experience, imagination and skill of the developer or user'.

Summary

The guidelines and heuristics which HCI have accrued appear to offer little explicit guidance to producing multimedia interfaces. Key issues in multimedia design are the delivery of information to support a task, and how this information will be processed by the user. Whilst the general principles of [Smith & Mosier 84] and the heuristics given by [Nielsen 92] will be true of any conventional interface in which an multimedia interface is embedded, little or no guidance is given to the problems of media selection, co-referencing and synchronisation. The comments of [Markus 92] are even more worrying; leaving multimedia interface design to the 'experience, imagination and skill of the developer or user' would seem to be one sure way of creating usability problems.

1.1.2 Instructional Design

Instructional design concerns how presentations should be designed for learning. Several instructional design methods have been proposed which could be useful as sources of guidelines for multimedia design. One of the earliest results concerns the use of different media for comprehension [Treichler 67]. It is suggested that people generally remember :

10% of what they read
20% of what they hear
30% of what they see
50% of what they see and hear

As educational technology moved from a behaviourist view to a one founded on cognition, such a result was however seen as far to speculative and simplistic.[Spencer 88] notes that these general findings 'may accord with many commonly held beliefs concerning the relationship between our senses and how we learn and remember, although they should be treated with caution, since no indication of their basis is given.'

In a review of media selection studies 15 years after [Treichler 67], [Clark 94] suggests that the problems of a non-theoretical approach still remains. [Clark 94] criticises educational technology studies of media selection for failing to realise that the issue was not one of which media to choose, but of the content of the media and how it supported particular instructional conditions : 'Media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck which delivers our groceries causes changes in nutrition.'

[Clark 94] notes that it is important to understand the role of the media's content in the process of comprehension : 'We need to ask whether there are other media or another set of media attributes that would yield similar learning gains'. In concluding [Clark 94] notes that 'cognitive efficiency' is the key issue in media choice, and that this may not be bound within a particular choice of media : 'It is important for instructional designers to know that there are a variety of treatments which will provide a desired learning goal. The designer can and must choose the less expensive and most cognitively efficient way to represent and deliver instruction.'

Several frameworks have been proposed in instructional design which have attempted to offer advice for the selection of media for educational tasks [Resier & Gagne 82], [Gagne & Briggs 88], [Anderson 83]. The issues that these model address are diverse and generally motivated by informing the needs of teachers; however they do also embody some general guidelines for multimedia design. It is noted by [Resier & Gagne 82] that most media selection models use the same underlying predictions, but address different levels of detail and different presentation media.

[Resier & Gagne 82] break the core issues of media selection into two issues : attributes of media, and learning task characteristics. A range of the model's predictions on attributes of media are summarised as follows :

> i) Visuals : The model distinguishes among various types of visuals, such as blackboards, pictures and also real objects. It is suggested that visual presentations will benefit poor readers.
>
> ii) Text : The model suggests text requires good reading skills; audio narration being favoured for poor readers.Other models indicate that print may be used in place of speech if different sounds do not need to be identified.
>
> iii) Sound : a distinction is made between verbal and non-verbal sound. Sound is suggested as valuable if the learning task requires recognition of the sounds themselves
>
> iv) Motion : the model differentiates types of motion, such as still visual, limited movement visuals, and full motion visuals. Motion is favoured if the learning task requires recognition or copying of the movements shown. It is also suggested that

motion should be used if the movements are unfamiliar or requires knowledge of manner of motion.

[Gagne & Briggs 88] stress that the value of particular media is related to the task for which learning is supporting. They provide guidelines which relate media selection to various learning outcomes. Learning out comes include :

i) Discrimination skills : make sensory discriminations
ii) Concrete concept skills : identify objects, classify objects
iii) Define concept skills : Verbatim learning, Substance learning
iv) Rule using skills : apply rules
vii) Problem solving : use knowledge of cause-effect relations

These are linked to different types of media and rated on their effectiveness for instruction. An example of the matrix used by [Gagne & Briggs 88] is shown below.

Functions	Media						
	Objects; Demonstration	Oral Communication	Printed Media	Still Pictures	Moving Pictures	Sound Movies	Teaching Machines
Presenting the stimulus	Yes	Limited	Limited	Yes	Yes	Yes	Yes
Directing attention and other activity	No	Yes	Yes	No	No	Yes	Yes
Providing a model of expected performance	Limited	Yes	Yes	Limited	Limited	Yes	Yes
Furnishing external prompts	Limited	Yes	Yes	Limited	Limited	Yes	Yes
Guiding thinking	No	Yes	Yes	No	No	Yes	Yes
Inducing transfer	Limited	Yes	Limited	Limited	Limited	Limited	Limited
Assessing attainments	No	Yes	Yes	No	No	Yes	Yes
Providing feedback	Limited	Yes	Yes	No	Limited	Yes	Yes

Figure 1.1 [Gagne & Briggs 88] matrix for media selection

A similar style of media selection model is proposed by [Anderson 83]. It is based on a flow chart, which asks various questions about learning objectives and supplies advice as to which media or combination of media is appropriate.

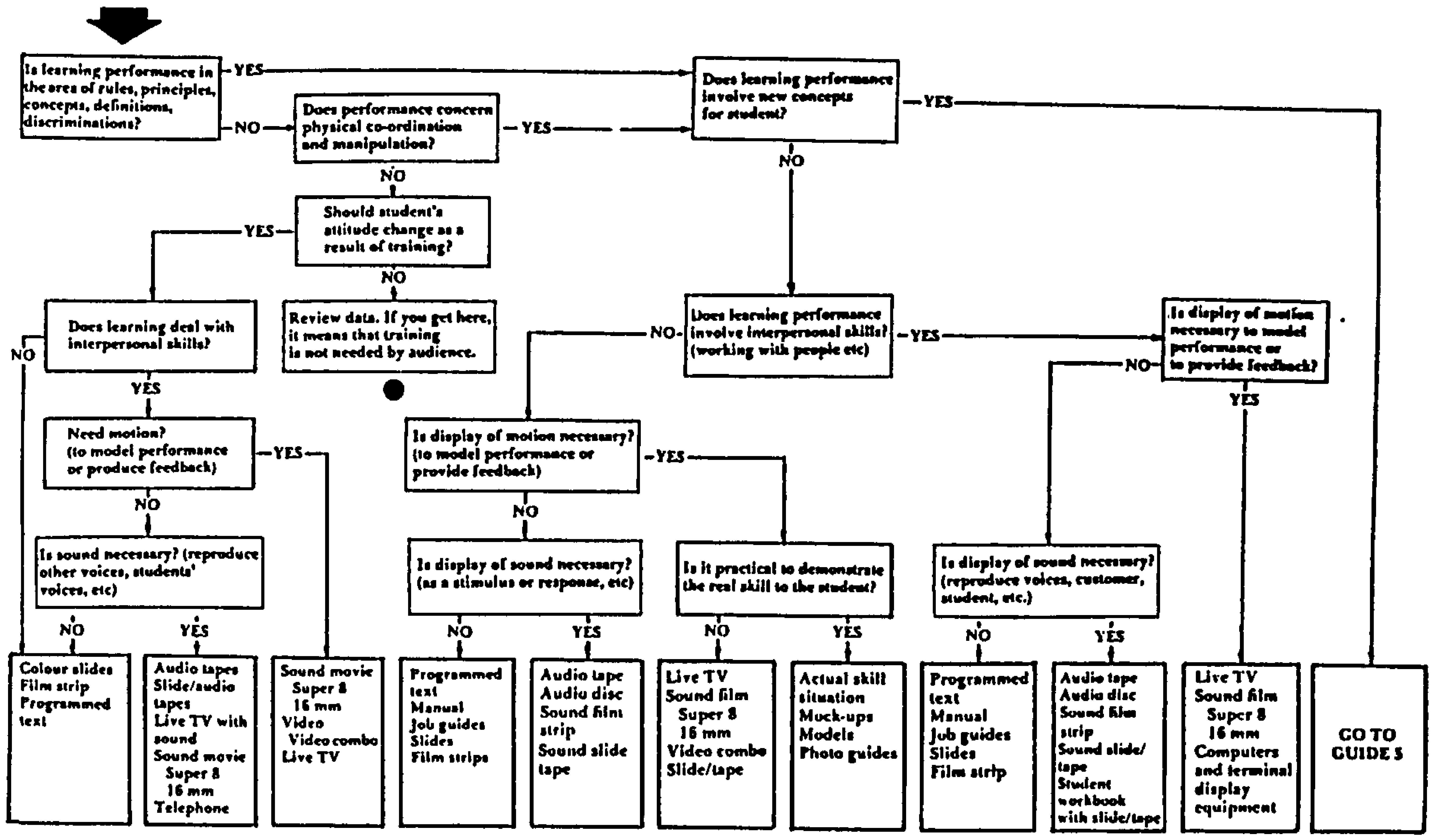


Figure 1.2 [Anderson 83] flow chart for media selection

A more detailed set of principles for instructional design of multimedia are provided by [Park & Hanaffin 93]. They summarise a range of educational issues in the use of multimedia. The following is a sample of the type of advice offered :

i) Present information across multiple, complementary symbols and formats
ii) Structure presentations and interactions to complement cognitive processes and reduce the complexity of the processing task
iii) Employ screen design and procedural conventions that require minimal cognitive resources, are familiar or can be readily understood, and are consonant with learning requirements
iv) Layer information to accommodate multiple levels of complexity and accommodate differences in prior knowledge
v) Organise lesson segments into mutually consistent idea units
vi) Differentiate concepts and principles through cosmetic amplification or repetition.

Several educational researchers [Salomon 79];[Kozma 91] have criticised instructional design models. [Salomon 79] attacks media selection models for failing to take into account the cognitive processes in comprehending the media. He suggests that the 'cognitive functions' of comprehension be related to media selection. 'We have such variables as pictures, line drawings, realistic visual, short and long texts, abstract and concrete ones, but rarely do we find a theoretical framework that relates them in nontrivial ways to cognitive functions they are expected to accomplish.' 'In sum, texts, pictures and there variable combinations affect learning not for their inherent attributes, but for the kinds of cognitive functions they can and do accomplish'.

[Kozma 91] suggests that studies of multimedia design by instructional designers will require a different method to that used for previous 'macro-level' studies of which media to use, noting further that the 'macro-level' approach has provided few useful results. 'Until now, the selection of media has been a macrolevel decision. That is the decision - should video be used or is audiotape sufficient - has been based on various instructional consideration in balance. This meta-question that has driven research on media for the past thirty years and has resulted in little understanding of learning with media.'

Instead, [Kozma 91] suggests a new 'micro-level' approach will be required to study multimedia design : '...media decisions for integrated multimedia will be micro-level decisions... The moment to moment selection of appropriate media can respond to specific learner needs and task requirements.'; 'A shift from macro-level to micro-level design decisions requires an understanding of the moment-by-moment collaboration between a particular learner and the medium.'

The types of issues [Kozma 91] raises for the 'micro-level' design relate to the way particular media support parts of a particular task rather than the task as a whole, and how the media will be processed by the learner to internally represent the task. [Kozma 91] concludes 'This moment-by-moment collaboration raises a different set of questions for the media researcher : what is the prior knowledge of a particular learner ? What symbol systems can best represent various components of the task domain ? How do these correspond with the way the learner represents the task ? How do they process various symbol systems together ? How can the medium process these in a way to support the learner ?'

Summary

The existing approaches to given in the instructional design literature are too high level and lack rationale. Whilst they seem to ask the right type of question, 'given a particular task need, what type of media should be used ?'; they offer little guidance as to how to combine and sequence media, or how to design the media themselves.

The criticisms of [Kozma 91] & [Salomon 79] suggest the need to form a model of the cognitive functions of multimedia and use it as a source for design guidelines. In particular, the [Kozma 91] notion of 'moment by moment design' implies a model which is able to address the use of different media to support particular parts of a task, and that the issues of media design, combination and sequencing should be considered.

1.1.3 Studies of Multimedia Comprehension

The area of multimedia design itself has received some investigation. At present, few empirical studies have been conducted which investigate how to design effective multimedia presentations.

[Mayer & Anderson 92] investigated the comprehension of a text and animation illustrating the operations of a piston (see figure 1.3). They claim that the addition of text enhanced understanding of the presentation as a whole, and they argued that to be effective the presentation must combine together both pictures and words into a unified whole. However little analysis is given as to what information is carried within the two media.

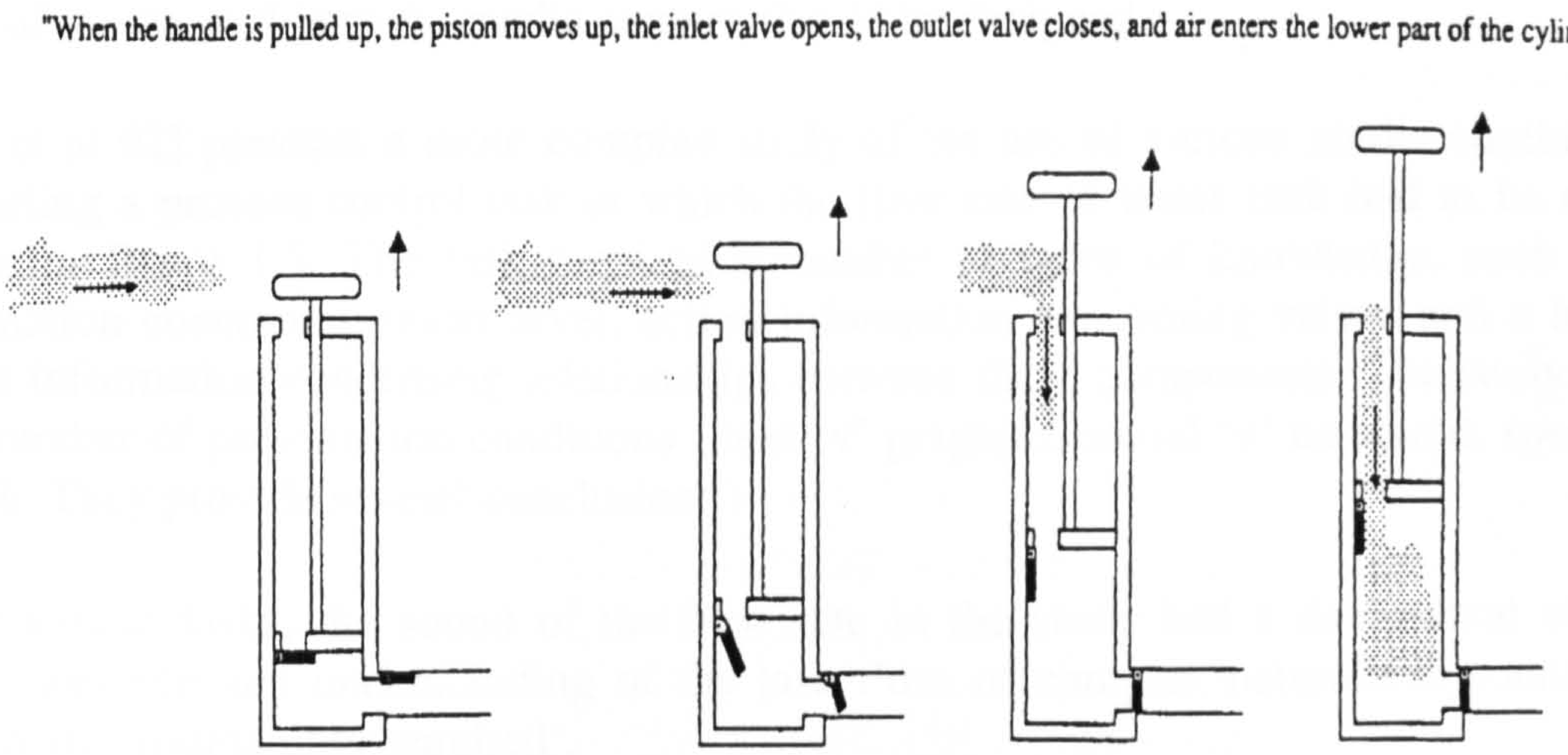


Figure 1.3 Materials used by [Mayer & Anderson 92]

One problem with both the studies of [Mayer & Anderson 92] is that it ignores the need to consider how well particular media fit to particular types of task. The study did not attempt to control for the media requirements of a particular type of task e.g. a task requiring spatial information 'v' one that needed procedural information. Thus it is unclear which task type was aided by which media or combination of media.

[Large et al 96] studied the use of either text, still image with text and animation with text in providing instruction either for a descriptive task (identifying parts of human heart), or procedural (how to find north using sun, shown in figure 1.4).

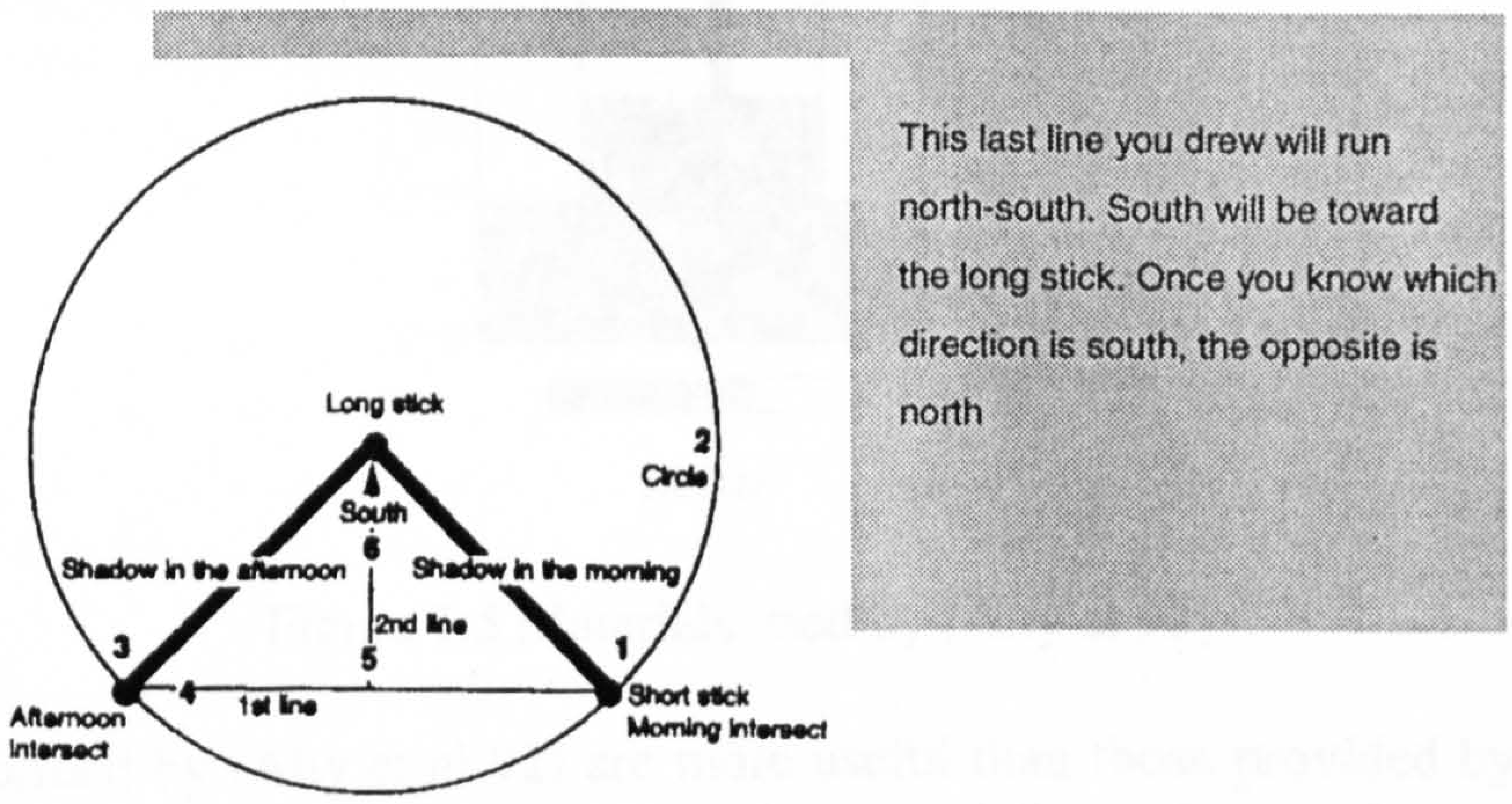


Figure 1.4 Materials used by [Large et al 96]

The results of the study showed that for a procedural task, animation aided comprehension; however for a descriptive task it was found that there was little effect for animation. [Large et al 96] concluded that 'a difficult text is not necessarily enhanced by an animation even when the animation is well-designed and highly relevant. Complexity is not removed by the addition of more media.'.

The results of [Large et al 96] acknowledge that the nature of the task will have an effect on the value of a particular media. However, as with [Mayer & Anderson 92], the studies are all or nothing in nature, failing to distinguish which parts of the task may benefit from particular types of media, and how the media content should be designed.

[Alty et al 92] presents a more complex study of the use of various media combinations in supporting a process control task in which the flow rate of water tank had to be controlled, shown in figure 1.5. The task required a number of types of knowledge, such as spatial information concerning water level, action information concerning valves and a heater, and causal information concerning relationships between these components. The study made use of a number of presentation conditions : text 'v' graphics, sound 'v' no sound, speech 'v' no speech. They provide several conclusions :

- for simple tasks, the sound of the flow rate in the study had a detrimental effect upon performance and understanding of the task :'use of complex naturalistic sound may not have discriminability required'.
- for more complex tasks, sound improved performance : 'the clear pattern that emerges .. is that operator performance with interfaces using sound was better for difficult tasks than for the easy ones'.
- speech warnings improved performance, but may be overly dominant in a presentation : 'speech did seem to improve performance during learning but is very intrusive'.
- graphical representations improved performance : 'for this particular task the graphical representation proved to be the most preferred. Thus our contention that this spatially orientated task would best be presented by graphics is born out by experiment.'

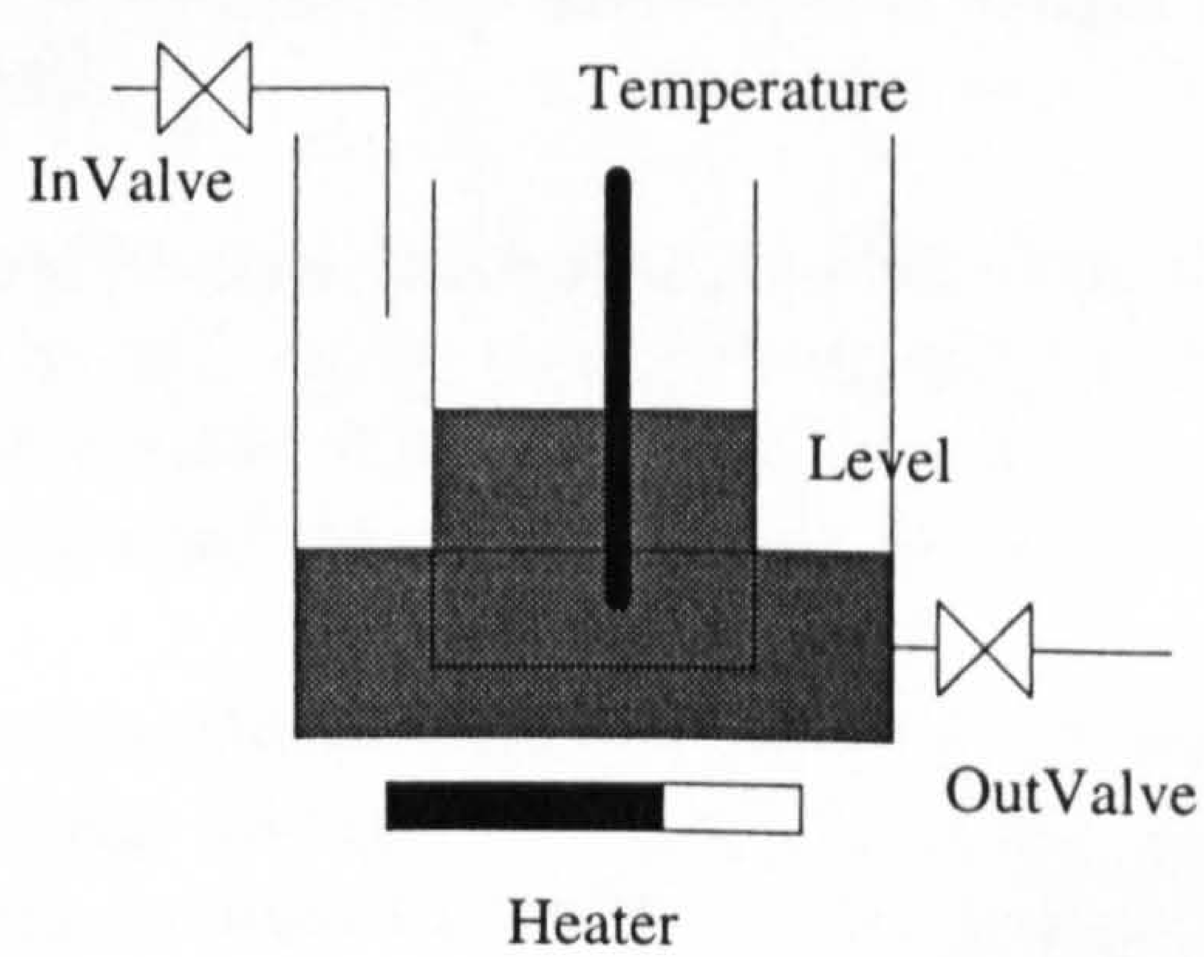


Figure 1.5 Materials used by [Alty et 92]

The study performed by [Alty et al 92] are more useful than those provided by [Large et al] and [Mayer & Anderson 92]. [Alty et al 92] explores the information requirements of the task and compares how the particular combinations of media used are related to them; and systematically made use of different conditions which varied the use of media for the particular information requirement. One problem with the [Alty et al 92] study is that its scoped to process control systems, rather than instructional tasks.

Summary

There is currently little empirical evidence of what may constitute good multimedia design. All studies reviewed offer some guidelines, but several suffer from two main problems. First, they deal only with all or nothing changes e.g. use animation 'v' text. This makes it difficult to distinguish if a particular combination of media may be useful.

Second, the studies often do not account for the type of task which is being presented, and explore how the media can be sequenced or combined to support it. The study performed by [Alty et al] shows a useful way to approach to these problems by systematically varying media.

1.1.4 Multimedia Guidelines

Several sets of multimedia design guidelines have been produced. A large number of 'cookbook' style books are available on multimedia presentation design. The following set of quotes is taken from a sample of ten, offering advice on the use of animation in multimedia :

i) Animation is visually attractive to the viewer 'Often video serves no purpose other than to offer an attractive display. It is not highly important that the viewer remember the content.' [Hodges & Sasnett 94]

ii) Animation should be used if other media are not sufficient. "add animation when still life won't get the message across.... resort to video only when other methods pale by comparison' [Vaughan 94]; 'Don't use motion video or animation or other multimedia techniques simply because you can, use them because they're the most effective tools for different situations' [Badget & Sandler 89]. [Fisher 96] suggests 'it is frustrating to try to describe in a series of still images an action which takes only a few seconds to watch.';

iii) Animation should be used when ever possible : 'the viewer must be able to digest each screen quickly and easily, use photographs and motion video where ever possible' [Badgett & Sandler 89]; 'If a picture is worth a thousand words, a video or animation is certainly worth many more.' [Burger 93]

iv) Still images may sometimes be better than animation. 'A photograph is often more evocative than moving footage' [Hoffos et al 91]; 'Essential information can be transferred from one medium to another - one striking still frame may convey the essence of a film clip.' [Hoffos et al 91]

v) Use animations if the process would be difficult to describe in language : 'a still image or video clip is the ideal support for textual information, especially for demonstrating a process.' [Fisher 96]; 'One reason that multimedia has the potential to be useful in education and training is that complex procedures can be conveyed more accurately in visual media than in verbal media.' [Davis & Marks 93]

vi) Avoid using video together with other visual media : 'when the viewer is intended to focus completely on the video in order to extract and remember its content, the video should not compete with other materials on the screen.' [Hodges & Sasnett 94]

vii) Animation can be used with other visual media : 'Remember too the video component of a given slide doesn't have to carry the majority of the information. Just as you can use a background with a still photograph, you can place video behind a text or graphic as a way of helping the viewer focus on a topic' [Badgett & Sandler 89]; 'the power of motion can transform a wall of dry text into a presentation which user's will approach enthusiastically.' [Hoffos et al 91]

It can be seen that these 'cookbook' guidelines are contradictory, and in many cases give very poor or misleading advice. Animation is suggested both as being the best media to consider, while also being a final option if other media can not be used. The specific conditions for choosing between using a still image or animation are unclear, some guidelines suggest still images may be better than animation, whilst others always favour animation. There is also a confusion between whether animation can be safely combined with other visual media or not; several guidelines note this as a serious problem, whilst two others seem to encourage it.

[MUMMS 96] provides a questionnaire for measuring the usability of multimedia systems. It is composed of a set of questions which the evaluator must agree or disagree. A sample is given below :

- Graphics, text and sound are put together in an attractive way.
- The quality of pictures and graphics is high
- This multimedia application has too much information.
- This multimedia application has a very attractive presentation
- I think this multimedia application could give me a headache

By attempting to be very general, the [MUMMS 96] fails to communicate many specific multimedia design principles. Many questions seem ad-hoc or vague, and it is unclear how the answers to such questions could aid in designing a more effective presentation.

[Heller 95] provides the most useful set of multimedia guidelines. These are presented as a checklist for multimedia evaluations, given in two parts. 'Technical measures' concern layout and design issues for text, sound, picture and video, whilst 'media integration' deals with how media are combined together. Checklist questions include rating the ‘use of colour’ between ‘not pleasing’ and ‘pleasing’, or rating use of ‘text with sound’ between ‘ineffective’ and ‘effective’.

The checklist produced by [Heller 95] is an improvement, in that it acknowledges the need for the design to combine media effectively, and includes the notion of synchronisation. However, [Heller 95] notes that 'what this taxonomy [of design issues] does not address is the design process and the order of use of different media', 'further this taxonomy provides only minimal guidance against the improper co-ordination of media, such as presenting a text on one topic and an image on another'.

[Heller] also fails to provide a rationale for the checklist. She notes in her conclusions 'the next step is to review the taxonomy in the light of various studies in the psychological and cognitive aspects of understanding of multimedia applications.' This suggests that [Heller 95] acknowledges the need for guidelines to be based on cognitive theory.

Summary

The current guidelines available for multimedia design are also poor. The [MUMMS 96] questions are very general, and would provide little real guidance.

[Heller 95] provides a useful set of checklist items. However, it falls short of providing a systematic method since it says little about selection or design of media, gives no details as to what effective synchronisation actually means, and does not provide any advice as to how fix the problems it diagnoses.

1.1.5 AI Generation of multimedia presentations

Research on the design of multimedia has been conducted within the field of Artificial Intelligence (AI) in attempting to automatically construct multimedia presentations. Several systems have been developed to explore these issues in the generation of multimedia documents. As these systems embed rules to produce multimedia presentations, they provide some implicit design guidelines.

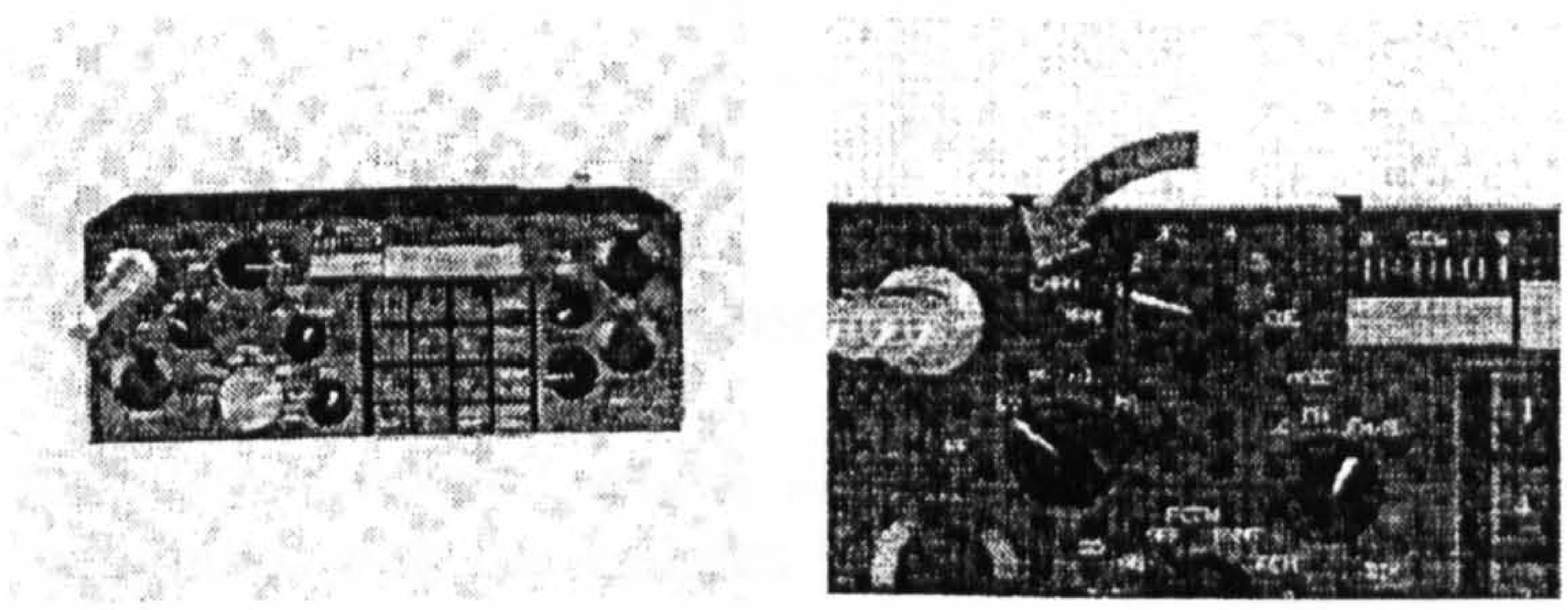


Figure 1.6 Screen shots from [Fiener & McKeown 91] showing COMET presentation for repairing a radio

[Feiner 85] produced the first system which automated the presentation of multimedia materials, called APEX. This system was designed to illustrate repair sequences for computer systems by providing annotated graphics. The only detail given by [Feiner 85] as to what rules may have been embedded within APEX is ' APEX incorporates the crude beginnings of a model for the creation of pictures that show actions, such as pushing, pulling or turning.'; 'in APEX we have attempted to eliminate unneeded detail in pictures while emphasising important features'.

[Feiner & McKeown 91] went on to produce COMET, a more complex system which generated both text and pictures to illustrate a repair task, shown in figure 1.6. Again, however, the descriptions of the COMET rule set are poorly documented; e.g. only following general principles are reported as being part of the planner :

i) Location and physical attributes - use graphics
ii) Abstract actions and connectives among action - text only
iii) Simple and compound actions - text and graphics

However, some details of how COMET produces images and animation are also provided by [Fierner & McKeown 91]. Cross references are proposed either when the user does not know the way an object is commonly described, or when the user requests the location of an object. In forming cross-references, it is suggested that 'graphical features are typically the easiest way to refer to textually and the most salient to the user'. However, little detail is given either of other guidelines embedded in the system, or on what principles they are based.

[Alty & Bergen 92] has developed a multimedia presentation planner for process control systems, called PROMISE. The system uses a taxonomy of media types specifying the available dimensions of each media type, whether the media is dynamic or static, has persistence, and has an information content.

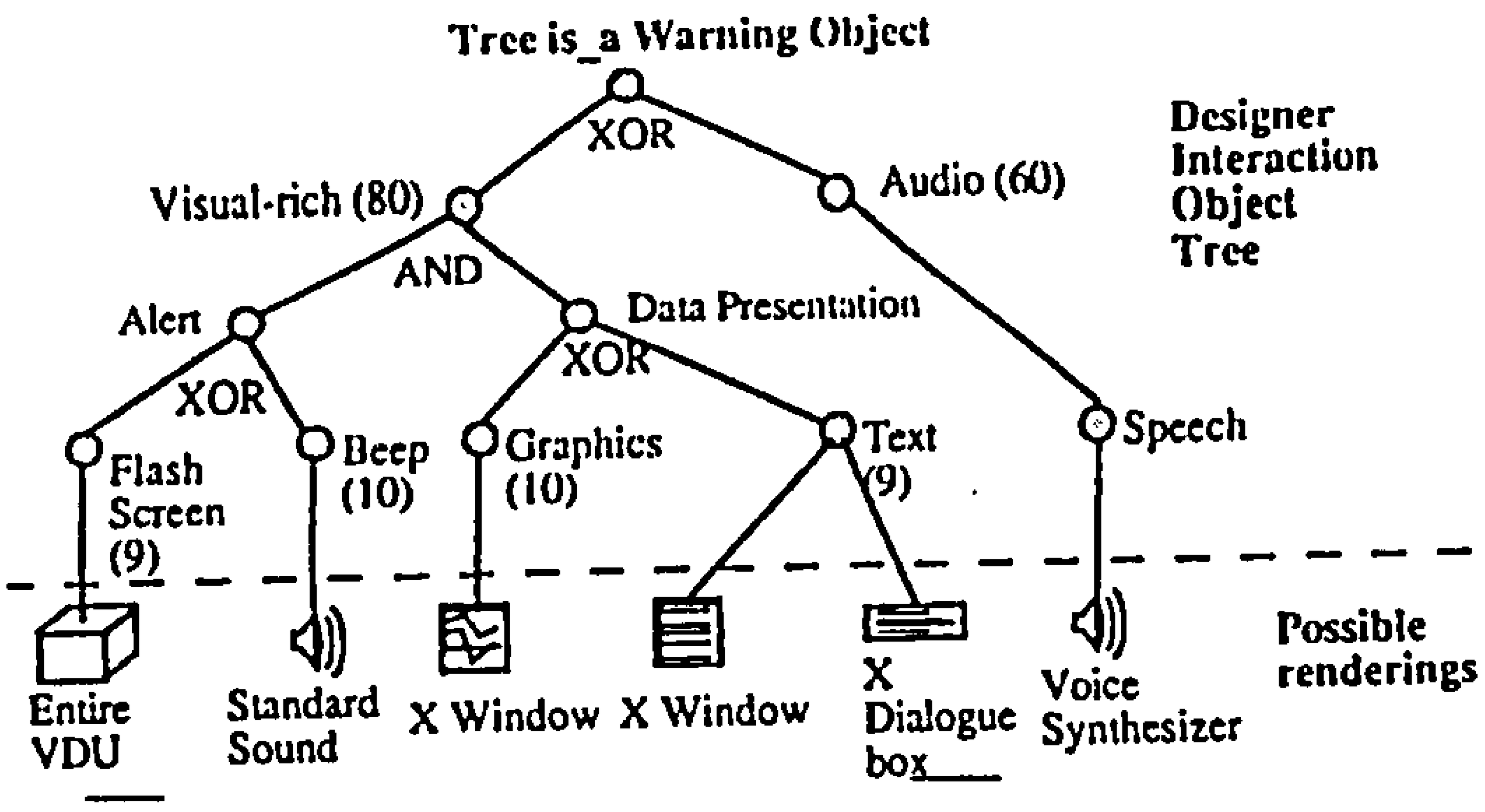


Figure 1.8 Decision tree used by [Alty & Bergen 92]

A set of decision trees are proposed which allow the various media to be synchronised together. These are weighted, with the weights being used in the planning process to decide which media to use. [Alty & Bergen 92] provides one of the few systems actually based on empirical study. The weights in [Alty& Bergen 92]'s decision tree were found by performing a number of studies on process control applications in which the choice of media were varied. The results were then used to inform the choice of weights used, see figure 1.8.

[Arens & Hovy 93] describe a planning system which attempts to select media based on the type of information to be communicated. They note that key issues for their planner are 'the characteristics of the media to be used', ' the nature of the information to be conveyed', 'the goals of the producer', and 'the characteristics of the perceiver'. Their system produces designs for graphs and text annotations. They summarise their rules as a table, part of which is shown in figure 1.6.

Medium	Dimension	Semantics	Endurance	Type	Detectability
Table	2D	2D	permanent	visual	low
Map	2D	>2D	permanent	visual	low
Picture	2D	infinite	permanent	visual	low
Sentence	1D	infinite	permanent	visual	low
Animation	2D	infinite	transient	visual	high
Graph	2D	1D	permanent	visual	low
Speech	1D	infinite	transient	aural	high

Figure 1.6 Media selection table used by [Arens & Hovy 93]

The rules given by [Arens & Hovy 93] are not backed up by any empirical study. Whilst the rules do take into account the viewer, such as whether the media will be 'detectable', they say little about the task, or the information content which needs to be communicated. A further problem with the rules is that they are of a very high granularity. It is unclear how a particular media selection is made based on criteria such as semantics, when both an image and text are suggested as having 'infinite meanings'.

One of the most detailed accounts of the workings of an multimedia planner is provided by [Andre & Rist 93]. The WIP system developed by [Andre & Rist 93] produces similar output to COMET. The following media selection rules are documented for WIP :

i) Prefer graphics over text for concrete information
ii) Spatial information, for speed use graphics, for accuracy use text
iii) For sequencing use text to express time change
iv) Express conditions with text

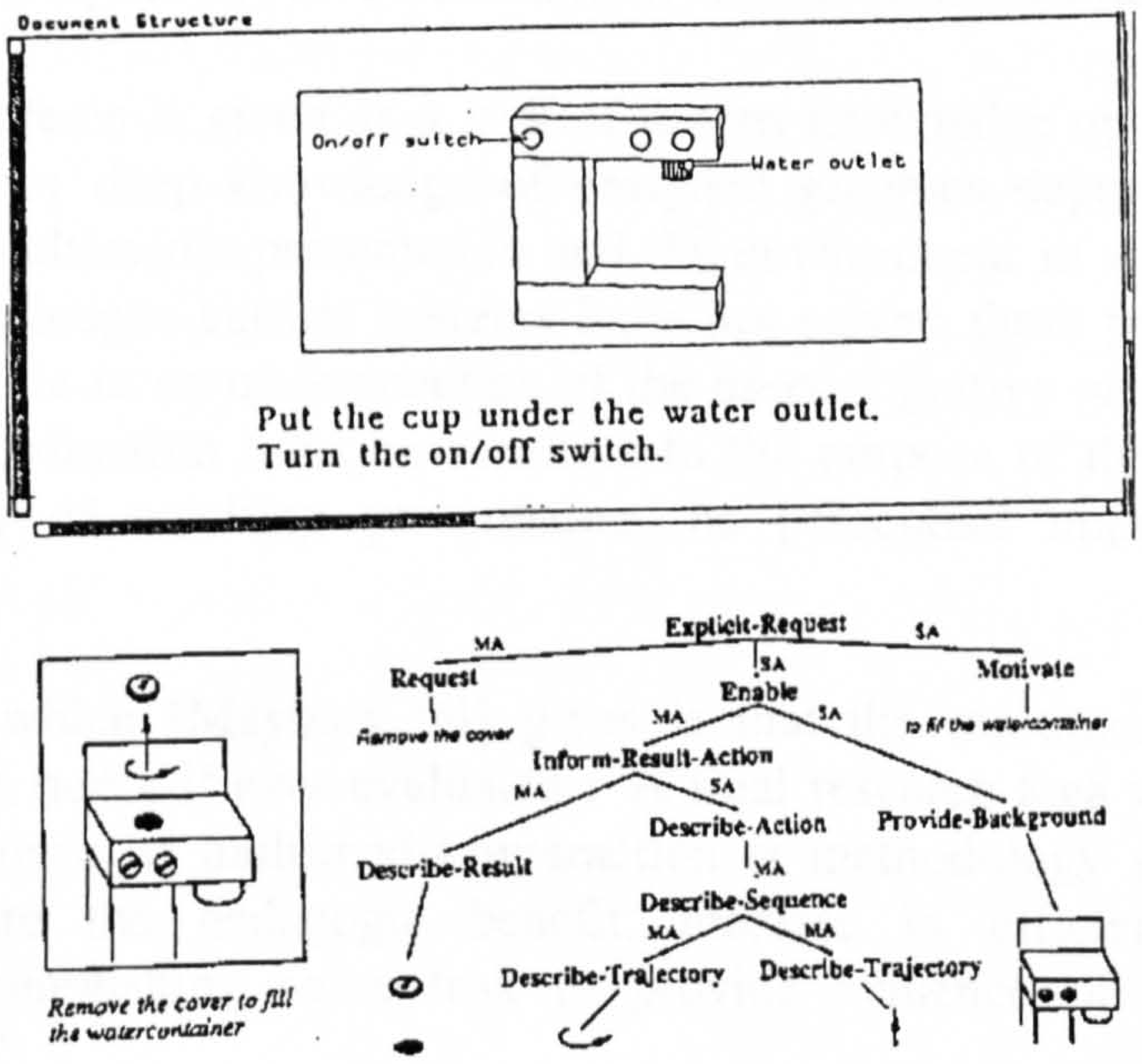


Figure 1.7 Screen shots from [Andre & Rist 93]

Another documented feature of the WIP system is its use of rhetorical structure theory in planning presentations, see figure 1.7. This provides a representation which the planner uses to support co-references between the text and picture in the presentation. [Andre & Rist 93] motivate this from earlier work on planning texts. A selection of the rhetorical acts documented are given below :

i) Attract attention : The text directs the addressee's attention to specific aspects of the picture-text e.g. 'Look at'
ii) Compare : two document parts provide a comparison between several concepts
iii) Elaborate : One part of the document provides further details about another part
iv) Enable : The picture-text provides additional information
v) Elucidate : One document part provides an explanation of another
vi) Label : A piece of text serves as a label

vii) Background : one part of the document establishes context for the other

The work of [Andre & Rist 93] shows the state of the art in multimedia presentation generation. The rhetorical acts are of particular interest, since they explain how the system deals with combing media in the presentation. The system does not however appear to deal with any form of dynamic presentations, so issues of synchronisation are not covered. Whilst WIP is documented far more explicitly than APEX or COMET, its value is reduced by the failure to anchor the choice of media selection guidelines and rhetorical relations within any theoretical framework.

Summary

In a review of work in automated multimedia generation, [Maybury 94] notes several difficulties. 'Many fundamental questions remain unanswered, including issues concerning the architectures and knowledge needed to support intelligent multimedia interaction, techniques for media integration and co-ordination, and methods for evaluation.'

One of the main problems is given as a lack of design knowledge upon which to build the systems : 'the need for deep knowledge of designed graphics depends at least upon the intended use of the multimedia presentation and the environment in which it is to be used'. [Maybury 94] acknowledges current systems have not solved these problems. He suggests that the solution may lie in an understanding of the users cognitive processes : 'Presentation composition and co-ordination must be sensitive to the purpose of the communication, the cognitive complexity of resulting presentation, its perceptual impact, consistency and ambiguity.'

The other criticism which [Maybury 94] gives is that the multimedia presentations the systems generated are rarely if ever evaluated : 'A final research area which can help foster process toward a science of multimedia interaction is methodology and evaluation'; 'it is important to measure the pedagogic benefit, increase in efficiency, or increase in effectiveness of accomplishing some task to provide evidence of the value added by additional machinery.'

1.2 Thesis Rationale & Organisation

The existing work reviewed suggests that whilst some guidelines for multimedia design do exist, they tend to be ad-hoc and poorly founded on theoretical or empirical studies. It is clear from the conclusions of [Kozma 91], [Salomon 79], [Maybury 94], [Heller 95] that the key issues in multimedia design are cognitive : how is a presentation understood by the user, and how can knowledge of the cognitive processes be used to improve design. [Kozma 91] and [Salomon 79] both conclude that the design of multimedia requires a clear theory of cognitive processes of attention and comprehension. [Heller 95] acknowledges this in the conclusions to her checklist for multimedia evaluation. [Maybury 94] also proposes that a solution to improve presentation design systems is to study the cognitive processes which are used to understand them.

Based on these arguments, the thesis has five main aims :

- To develop multimedia guidelines and principles from a synthesis of the literature and experimental study. The thesis will resolve contradictions in the current multimedia design studies by reference to psychological studies and models, and empricial evidence.

- To investigate the cognitive process of attention and comprehension by experimental studies on multimedia. None of the studies reviewed explored the available literature in cognitive and educational psychology pertinent to the multimedia design problem.

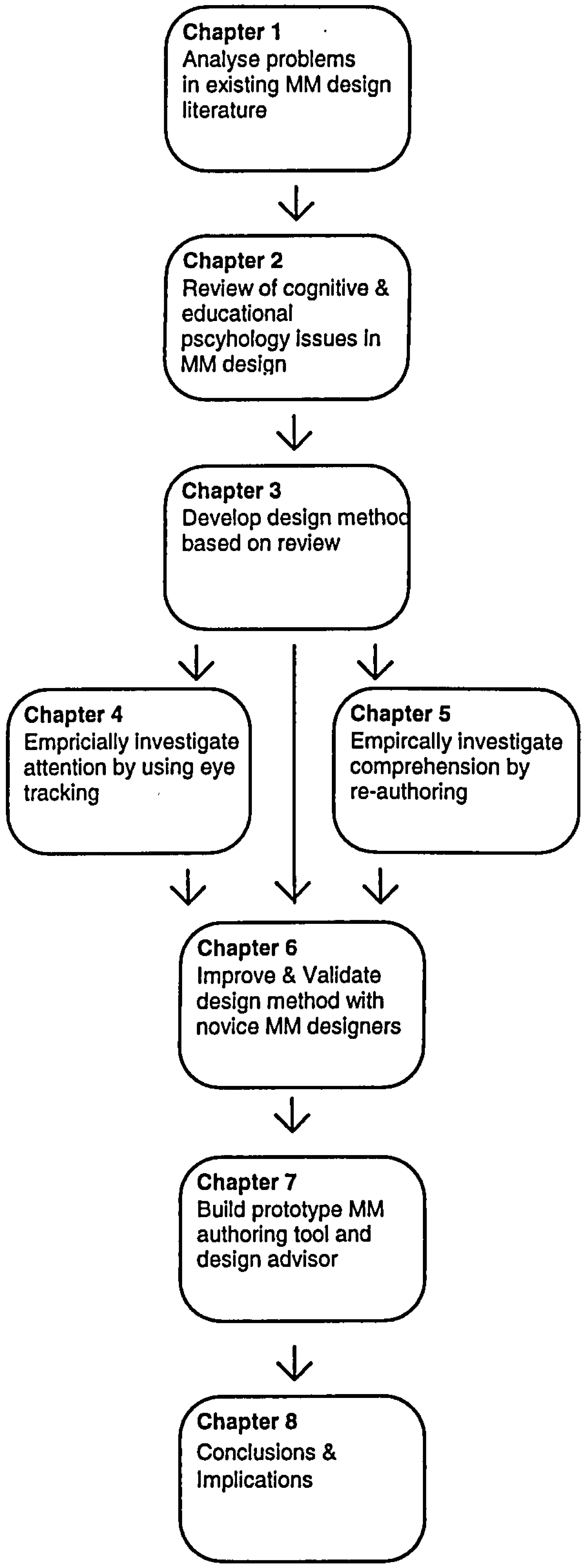
- Create a method for multimedia design incorporating the principles and guidelines. In order to make the work tractable for use by real world designers, in must be motivated by a method.

- Validate the method and tool by case studies with industrial practitioners. If the method is to have value, it must be usable by the designers in the real world in assisting them in producing multimedia presentations.

- Development of a tool to deliver the method's advice within an authoring environment. Since multimedia design work is largely carried out within tool based environments, it is important to explore how the method could be delivered via tooling.

In order to scope the work to a more manageable area, the thesis will deal with only the issues relating to presentation design. Thus the focus of this thesis is on delivering expository content to the viewer, and not how an interactive sequence should be designed or structured.

The remaining seven chapters describe how the aims of the thesis were met. Chapter 2 provides a framework of cognitive issues in attention and comprehension of multimedia. This work is structured around existing studies of the use of speech, text, pictures and animation in cognitive and educational psychology. It begins by discussing perception and attention to auditory and visual information. It then considers how verbal and visual may be combined in a presentation. A conceptual framework is then developed to structure a survey of comprehension of different media and how media can be designed effectively.



Chapter 3 presents a paper based design method built upon the cognitive framework given in chapter 2. It focuses on providing guidelines for media selection, within media design, attentional design of image, animation, text and speech media. A case study is used to illustrate the method steps, showing how a task model is used to choose different information types, how media are selected, how within media design decisions are taken, and how attentional design is considered.

Chapter 4 and 5 report empirical studies conducted to validate some of the method guidelines. Rather than attempt to test each guideline in a factor by factor study, which would have proved impossibly complex and time consuming due to the number of guideline combinations possible, the studies were exploratory. The studies attempted to find examples of good and bad design in an existing commercial presentation, then compared the results with the method guidelines.

Chapter 4 presents a set of eye tracking studies. The studies were performed to understand how people view an multimedia presentation. Three presentations were considered, with the eye track results used to suggest possible design problems. The results are used to motivate attentional guidelines for multimedia design.

Chapter 5 represents three further studies on the materials used in the eye tracking study. The first study explores how well subjects understood the original multimedia presentations, comparing experts with novices. The results are again used to suggest possible problems. The commercial presentations are then re-authored. Two further studies are then reported. The first compared the original presentations with the re-authored version. The second then validated the changes made in re-authoring.

Chapter 6 presents a validation exercise on the use of the paper based design method. Novice designers were given seminars on how to use the method, and then used it design a multimedia presentation. It shows how the method helped to inform them in preparing storyboards for a presentation.

Chapter 7 addresses how the paper based design method was used to within a design support tool. The tools architecture and interface are discussed. The use of critics and expert system based wizard advisors to provide design guidelines is summarised. A case study example is then used to show how the tool provides support and may help prevent design problems.

Chapter 8 summarises this research and proposes future directions. Implications for the use of the method, and for further development of the design support tool are discussed.

1.3 Research Contributions

- A model of attention and comprehension of multimedia is proposed. It summarises a range of existing studies on speech, text, pictures and animation.

- A systematic method for multimedia design is developed. Guidelines are provided for media selection, within and between media design and attentional design.

- The importance of attentional design in multimedia is explored. Eye tracking studies of multimedia presentations indicated problems in attention and suggested design solutions.

- The importance of within media design is investigated. Studies show how design changes can improve user's understanding of multimedia presentations.

- The need to provide design support within authoring tools is suggested. A demonstrator authoring tool shows how future multimedia tools may be used to aid the user in making design decisions. Critics and expert system based 'wizards' are implemented as ways of providing this assistance.

- The value of a multimedia design method was validated by testing it with novice designers.

Chapter 2

Framework for Multimedia Attention and Comprehension

2.1 Introduction

The process by which a user comprehends a multimedia presentation is a cognitive one. Thus, in order to produce a set of guidelines for the design of multimedia presentations, a model of these cognitive processes is required. This chapter is made up of a survey of existing educational, instructional and cognitive psychology studies concerning perception, attention, short term memory, working memory, and comprehension of different media and combinations of media. These studies are used to propose a framework for cognitive issues in the design of multimedia.

The structure of the framework is based on models of cognitive information processing, such as [Barnard 85], [Baddeley 86] and [Miller & Johnson-Laird 76]. These models suggest that several processes are required to move from images and language in the world to meaning in the viewer. These models provide general architectures for how meaning is processed from the world. They all contain several components :

> (i) Perception, Attention & Short Term Memory, concerning how information is initially extracted by attention and buffered whilst it awaits processing
> (ii) Working Memory, how information is chunked together and cross-referenced.
> (iii) Comprehension, how meaning is extracted for the information.

These general architectures are refined in this chapter to address issues specific to multimedia design, rather than human cognition in general. They are placed within a framework for MM design which is shown in figure 2.1. It has several components. The role of each component in the framework and its implications for multimedia design is discussed briefly below, and in more detail in the following sections, followed by a discussion of their value for design.

The framework first covers perception and attention to different media, and combinations of media. This section surveys literature on perception, attention and short term memory of auditory and visual information. The review is structured around perception and attention to different media types : speech, sound, text, image and animation, and to combinations of language and image media. The issues for multimedia design concern 'attentional design', how elements of the presentation should be designed to make them more salient, or 'emphasise' them; and what restrictions short term memory (STM) places on how auditory and visual media elements can be presented together within a presentation.

The second section deals with how working memory is used to combine verbal and visual media together. This builds upon the first section by comparing how attention and working memory can be best used to bring together speech, text, image and animation to form a single whole. 'Contact points' are introduced as ways in which visual and verbal media are integrated by co-reference. Issues for multimedia design include how to bring combine media to form a single whole, and how to effectively design 'contact points'.

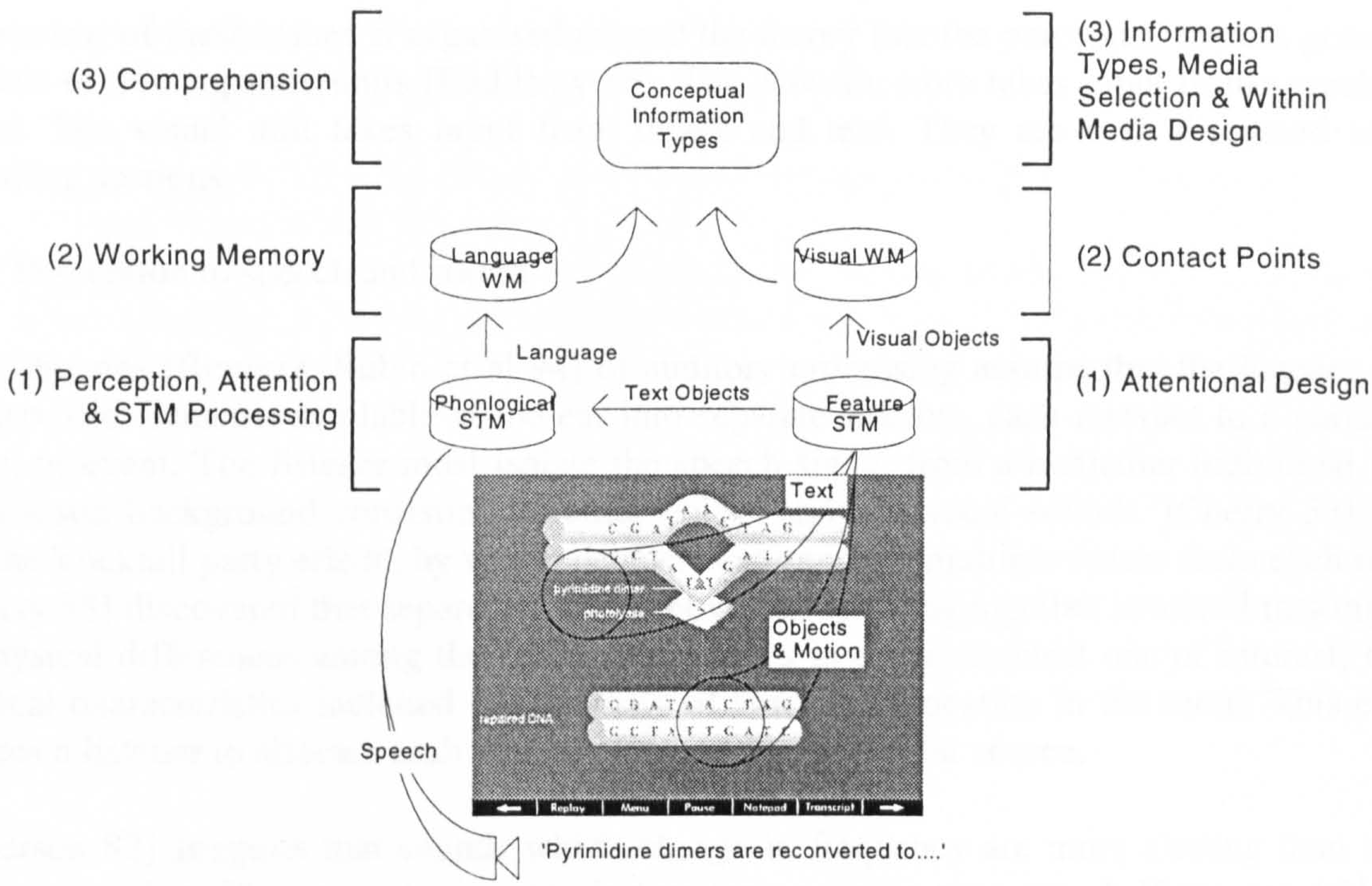


Figure 2.1 Overview of cognitive framework (left-side) and related multimedia design issues (right-side)

Finally, comprehension of multimedia is addressed by surveying psychological studies which have investigated which media are suitable for particular types of content, and how the media should be designed to convey their content. The review is structured around a set of 'information types' based on work by [Miller & Johnson-Laird 76] & [Jackendoff 83]. The information types abstractly specify the type of content to be communicated. The review groups together studies concerning the effectiveness of language, image and animation in supporting each information type. It then addresses how media should be designed to support comprehension of the information type. The design issues concern 'media selection', which media is best to support comprehension of a particular information type, and 'within media design', how should the media be designed to aid comprehension of a particular information type.

2.2 Perception, Attention & STM processing

The key issues for multimedia design concern attentional design and restrictions STM capacity place on processing an multimedia presentation. These include :

- How to plan the attentional thread to guide the user's viewing and reading sequence

- How to design effects and combinations of media to make important information salient and combine information presented via language and image based media.

- How to determine timing and synchronisation so that the user has sufficient time to assimilate the content

The review of these issues is organised around the theory that the perceptual system generally consists of two separate units [Baddeley 86]. The acoustic store takes input from speech and sound. The visual unit takes input from image and text. They are each discussed in the following sections.

2.2.1 Perception to speech and sound

Most theories [Remeze, Rubin et al 94] of auditory processing assume that the listener sorts the acoustic elements available at the ear into separate streams, each relevant to a particular object or event. The listener must isolate the speech signal from a particular individual from an acoustic background consisting of other voices and non-vocal sounds. [Cherry 53] calls this the 'cocktail party effect', by which listeners can separate multiple voices from each other. [Cherry 53] discovered that separating different voices speaking together involved making use of physical differences among the auditory messages in order to select one of interest; these physical characteristics included the intensity of voice, and location in the room. This effect enables a listener to allocate each stream of sound to a particular source.

[Patterson 82] suggests that sounds which change in frequency are more alerting than those which are stable. Thus speech is more alerting than a single tone sound. [Patterson 82] also notes that the separation suggested by [Cherry 53] is found only for speech; acoustic elements of similar frequency are attributed to the same source (in terms of change and of temporality). Thus, noise and non-speech sound tends to be grouped together and is separated from speech, while sounds of similar frequency tend to be grouped together and not distinguished.

Sound and speech are linked to a source in world, however spatial localisation is generally poor [Perrot 91]. [Mountford & Gaver 89] note that sound can give information on physical actions (e.g. breaking), physical transitions (e.g. filling up glass) and spatial location.

2.2.2 Speech

Short term memory for speech is based on echoic memory [Baddeley 86]; which provides a store of surface sound for around 20 seconds. The echoic store is uni-directional, sound can only be 'played' out in a forward direction. Echoic memory acts as a buffer; from which speech is recoded into morpholexical working memory. This holds around two simple sentences which will be overwritten as further speech is processed. Morpholexical STM allows limited re-instatement of speech information which is no longer available from auditory perception. This can assist with identification of complex or difficult words.

[Baddeley & Salame 82] have found that background or 'unattended' speech may gain entry and also cause overwrite within the morpholexical WM. This was found to be the case even if the speech was not understood e.g. in foreign language. However, it was also found that sound and noise do not interfere with speech in morpholexical STM.

Speech is made up of approximately 40 phonemes in spoken English [Nickerson 80]. [Rubin 80] provides a summary of speech identification. He notes that speakers produce words at around 150 per minute (10 - 12 phonemes a second) and notes that processing must be based on chunks rather than individual phonemes. [Rubin 80] also suggests that the way in which speech is produced will also effect meaning, noting that 'spoken language has salient use of stress, intonation and other prosodic features'. The following effects are suggested :

- Temporal characteristics : pauses and changes in speed often provide clues for the grouping of words into phrases.
- Speed of production : More quickly spoken words indicate something of less importance.
- Stress : stressing particular phrases may separate the as important or new information
- Prosody : the rhythm in which words are spoken can allow illocution, pauses and lists to be identified

Some evidence for [Rubins 80] claims is given by studies performed by [Palmeri, Goldinger & Pisoni 93]. They investigated the effect that voice attributes had upon recognition of speech. Their study compared recognition of words from a similar voice, or from a new voice. It was found that subjects recognised specific voices quite accurately. They concluded that 'these results suggest that detailed information about a talker's voice is retained in ... representations of spoken words.'

Summary

The following properties of the speech and sound are relevant to multimedia perception :

- Speech and sound are difficult to ignore, speech in general can not be 'unattended'; unnecessary speech or sound may distract attention.

- Speech will be separated from sound by perception. Only a single strand of speech or sound can be in focus at once.

- Multiple strands of speech or sound may interfere with each other and distract attention.

- Voice, speed, pauses, stress and prosody may allow particular words or phrases to be emphasised in speech.

- Use sound as a warning or to initially gain attention

2.3 Perception to Visual Media

By default, perceptual system uses auditory information as a secondary source of information. The visual system is used as a default. This is termed visual dominance by [Posner 76], who conducted a series of studies comparing the response to auditory stimuli (single tone) against visual stimuli (flashing light). [Posner 76] proposes the following rules :

> i) Visual stimuli are not as automatically alerting as auditory stimuli. This is because visual stimuli require active attention, auditory stimuli do not.
>
> ii) To compensate for the low alerting capability of visual stimuli, an attentional bias is given to the visual system when a choice is to be made between auditory and visual.

[Posner 76] concludes that the visual system is the default taken for processing : 'This interpretation implies that in conflict situations, vision will usually be dominant.' However, it will be 'alerted' by the auditory system : auditory information is processed without attention and will then signal the visual attention process, speeding up its response.

2.3.1 Perception

The fovea lies at the centre of the eye and has a greater density of receptors, giving higher visual acuity. It covers roughly 2 degrees of visual angle [Nelson & Loftus 80]. [Rayner 81] provides a tutorial review of visual perception. He suggests that the direction in which the fovea points generally gives the area which is fixated. Outside of this area lies the parafovea, which covers a wider area of approximately 5 degrees of visual angle, but at a lower level of detail, having a lower density of receptors (see figure 2.3.1). Very little useful information can be extracted outside of the parafovea. Because the fovea covers such as small area, the point of fixation must be moved for the eye to process new visual information. Repositioning of the eye is termed a saccade. During a saccade, no visual information is extracted.

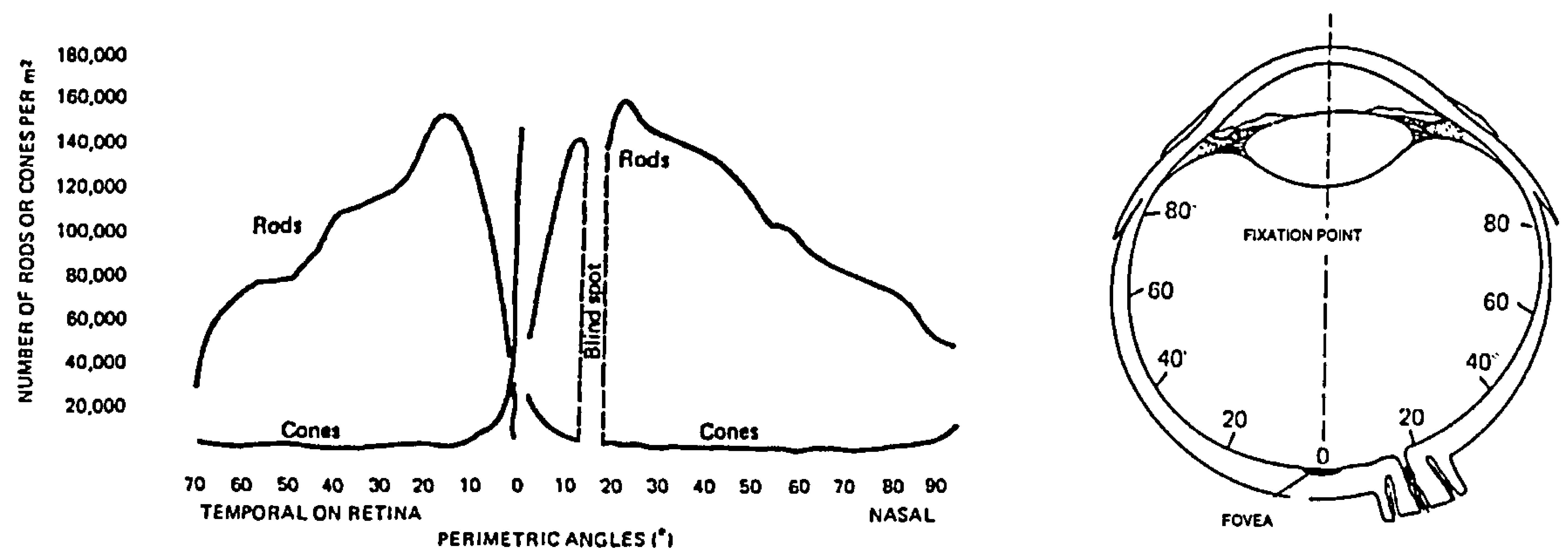


Figure 2.3.1 Fixation point of the eye and fovea

Attention can be modelled as a spotlight moving over a bitmap made up from the image [La Berge 83]: the area of the bitmap being 'illuminated' by the spotlight can be processed further into visual STM [Johnston & Dark 86], [Treismann 88]. The spotlight is independent of the point of fixation and may operate upon either the fovea or parafovea [Henderson et al 89]. [Eriksen & St James 86] note that there would be little point in moving attention outside of the parafovea area, 'where so little detail resolution is provided by the visual sense organ that the concentration of processing capacity is essentially wasted.'

2.3.2 Text

[McConkie & Zola 82] suggest that attention processes text at the word level, rather than sentence or page (see figure 2.3.2 for an eye track). This may be due to a word being close to the maximum size or 6-8 letters that can be processed within the fovea at any one time (letters require roughly 1/5 of a degree of visual angle, whilst lines require 5-20 degrees).

[Loftus 83] suggests that saccades tend to move across the page in word size jumps; the fixation point is usually at the centre of the word and changes 4 or 5 time a second (200 msec). However, not all words are fixated upon e.g. the initial and last word on a line. Longer or more complex words tend to be fixed on for longer than shorter or less complex words. Average reading speeds for two lines of text (80 characters) are around 6 seconds. [Loftus 83] also notes that whilst reading follows a predominate left-right, top-bottom order, occasionally words will be re-fixated. They also propose that word identification requires that both the fovea and attentional spotlight to be focused upon a particular word. Word identification makes use of schema, which provide templates against which the word as a whole, or letters with the word are recognised [Treisman & Souther 86].

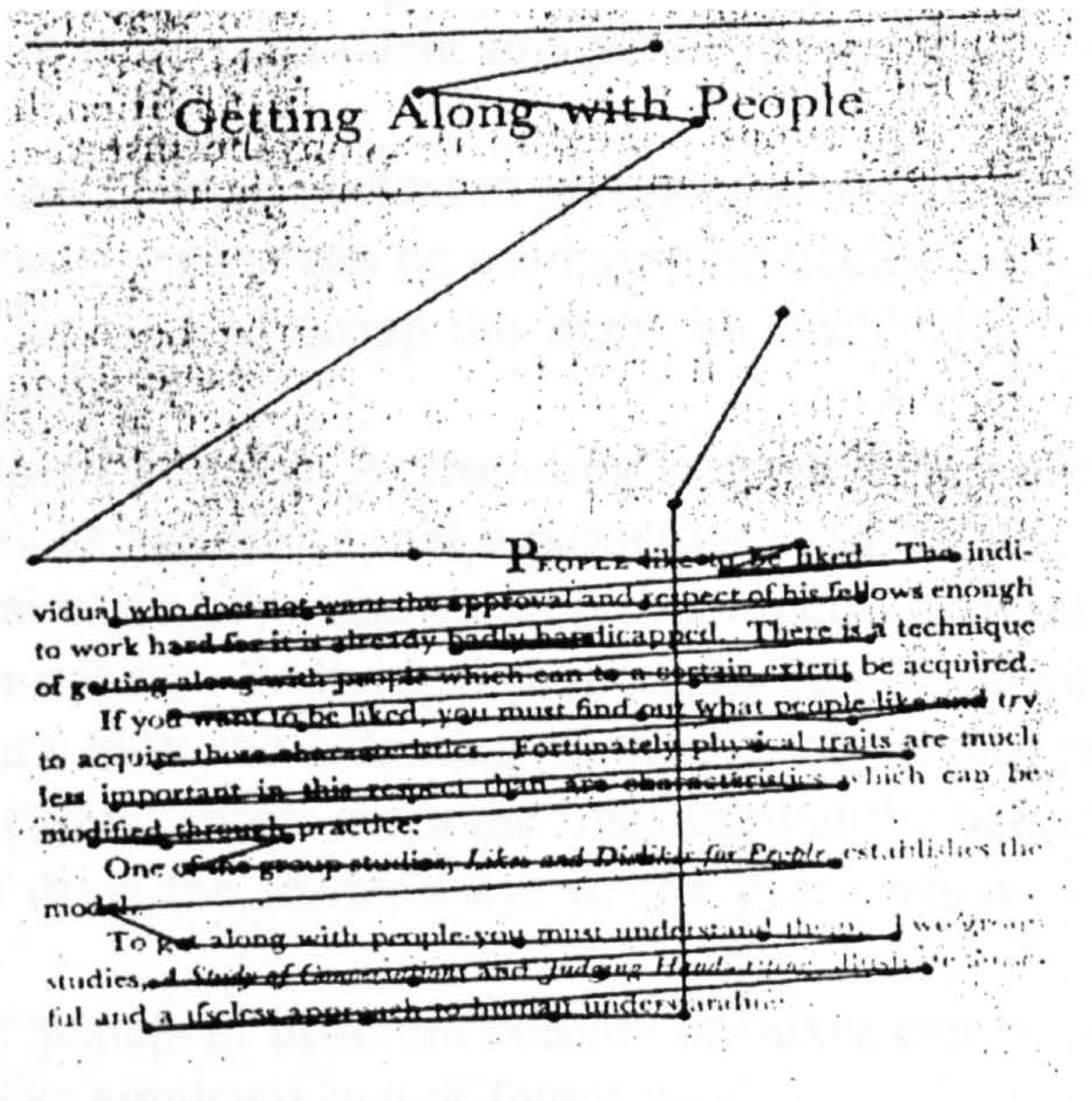


Figure 2.3.2 Eye track of reading [Yarbus]

Several studies have been conducted upon how typeface and text design effect attention and recognition. [Lewis & Walker 89] suggest that physical typeface information may be attended to. They note that certain type styles, such as type face, weight or font size, italics, underlining are attended to and thus become part of meaning of the text, or provide a 'typeface personality' upon the text. A study was performed in which the typeface was changed to disagree with the content of the text e.g. unimportant text was made bold. They found that mis-use of type style reduced comprehension. The conclude that 'despite the primarily alphabetic nature of English, it possesses an element of pictorial reference which is contingent on typography.'

[Glyn, Britton & Tillman 85] provide a review of the use of typographical cues in educational text. They suggest that cues serve a number of purposes : '1) to set off important items and enhance their recall; 2) to set of relatively large chunks of supporting information from the main body of text'. They note that using italics, bold font, colour, boxes, or separating text with white space all serve to draw attention : 'typographical cueing systems are nonverbal devices for attracting and focusing the readers attention.'. [Glyn, Britton & Tillman 85] summarise empirical results which suggest that the use of these techniques improves recall of the parts of text effected : ' the provision of a cue such as underlining increases the learning of cued information'. They also conclude that this may be at the expense of other uncued information and suggest that to separate large bodies of text, white space or a box should be used, since using too much underlining or bold font will reduce readability.

More general typographic guidelines can be found in design guides for magazine design and desktop publishing. The following guidelines and motivations are provided by [Hartley 78] :

- Use of bold : 'bold is the easiest and most effective means of giving emphasis, but it may confuse or limit the variety of bold subheadings'
- Italic : 'italic will change the pace sufficiently and less insistent than bold but it may be designated for unusual words'
- Capitals : 'give very strong emphasis so they should be used only if a few individual words need emphasis', 'word images built from capital letters contain less distinctive visual information than do equivalent images of the same typesize, so recognition may be less immediate'
- Space : 'line space can be as important in giving emphasis as size and weight, more space gives greater emphasis', 'space can be used systematically, not only to separate the items from one another, but also to group the items hierarchically by employing one, two or four units of line spacing',
- Indentation : 'identing lines, that is preceding them with a space, can be a useful device for giving emphasis or indicating that a particular piece of information is separate from another', 'text can be indented to show that it is somehow different from the main copy'
- Bullet points and numbers : 'individual paragraphs can be picked out by using a marker such as a number or a blob and indenting all the rest of the paragraph'
- Initial letters : 'big capital letters are sometimes used at the beginning of paragraphs. their purpose is both to draw the reader's eye to the point where the text begins and to be decorative'.
- Panels and colour : 'panels of different colours or boxes can be used to separate different sorts of text or to give emphasis in a different way'

Summary

- Reading time must be allowed before attention can be directed elsewhere. Longer, complex and unfamiliar words will take longer to process.

- Formatting can direct attention to the required part of text e.g. paragraphs and titles provide entry points for fixation to searching text; bullets and lists guide attention.

- If a particular word or clause of text is important make it stand out e.g. use highlighting, bold or large fonts or underlining

2.3.3 Image

Whilst the attentional processing of a text is an ordered affair, images are not processed in any particular order. In general, the first pass of fixations upon an image will make use of the both the parafovea and fovea to process as large an area as possible [Rayner 80]. Because of the large area attended to, the information provided for object identification will be general. The fixations at this stage will range over large areas of the scene [Nelson & Loftus 80].

The first pass of fixations are used to provide a general information on surface and form, but with little detailed information extracted. [Venturino & Gagnon 92]; [Palmer 75] have both demonstrated that the first pass of fixations will give gross details, see figure 2.3.4. [Palmer

75] showed this phenomena with a 'fruit face', the first pass would extract a face, whilst subsequent passes would give details of the fruit. [Venturino & Gagnon 92] extended this finding to real world scenes by testing trade-offs for different levels of identification if attention was focused over a wide or narrow area. They conclude that 'visualising a large area such as a scene may make the details of the smaller elements less accessible'.

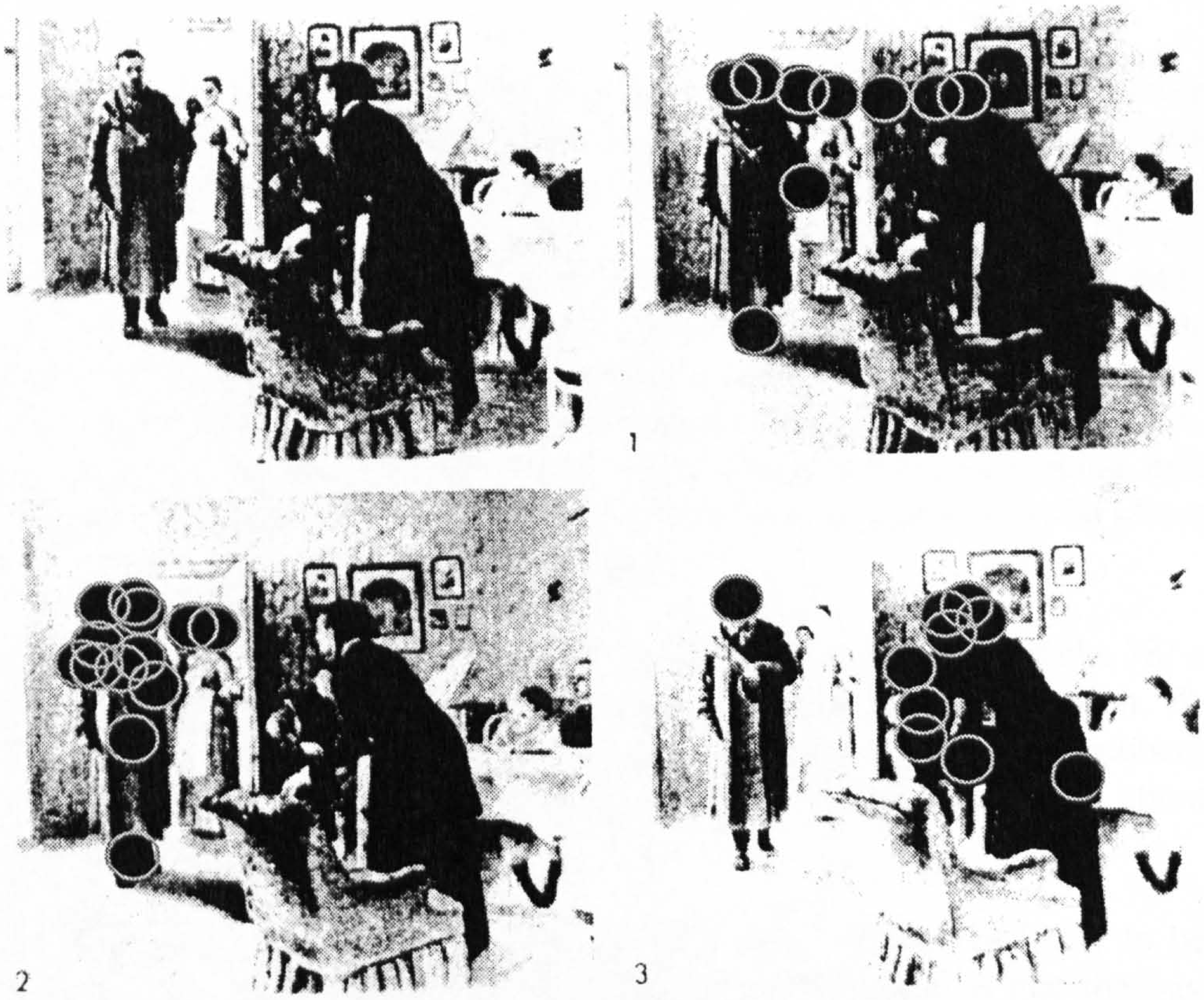


Figure 2.3.3 Eye track of phases of fixation to scene [Yarbus 67] : a first pass of fixations can be seen in frame 1, followed by more detailed fixations in frame 2 and 3 to people in the scene.

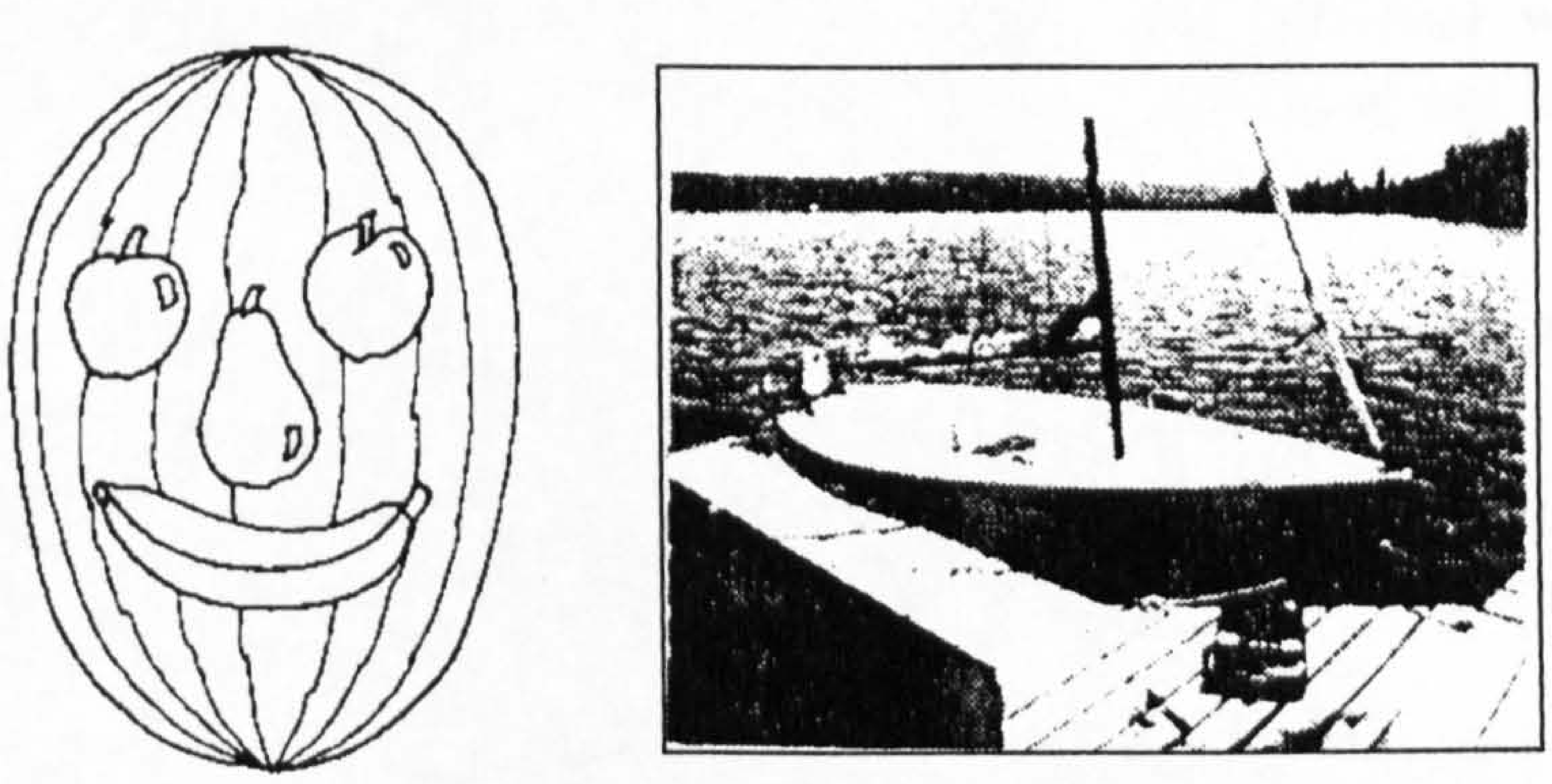


Figure 2.3.4 Fruit face [Palmer 75] showing that objects are first recognised as a scene; test scene used by [Venturino & Gagnon 92]

Following the first pass of fixations, a second phase of more detailed inspections begins. This uses more directed saccades to position the fovea over smaller more localised areas of the scene [Rayner 80]. [Yarbus 67] provides an eye track which shows the first few fixations scan across the scene, then subsequently cluster at particular points of interest (figure 2.3.3). On average, the eye saccades three to four times a second during scene processing, and will usually move between 2 and 6 degrees from its previous fixation point [Rayner & Pollatsek 92].

[Hochberg & Brooks 78] investigated how changing the length of time images were available to be viewed for effected memory for image detail. By studying the number of fixations given to an image, it is suggested that the scene or animation is initially scanned for general information, then for more specific detail : 'a fast component brings the eye to those peripherally visible regions that promise to be informative, and a more sustained component directs the eye to obtain more detailed information about the main components that have already been located'. They note that the speed at which a new image is shown or 'visual momentum' should be set at a point which allowed for both processes to take place : removing the image prior to this resulted in more detailed visual information being missed. These results suggest that attention to images is a serial process, and time must be allowed for it to take place before changing or removing the image.

The results of [Venturino & Gagnon 92]; [Palmer 75]; [Hochberg & Brooks 78] suggest that images are first processed as a whole, giving a general scene level information. They are then fixated upon object by object, to produce a more detailed description. [Hochberg & Brooks 78] note that this process takes time, and that scenes must be left on view for attention to scan at the gross then narrow levels.

Several studies provide useful results about what kind of details are likely to be picked up during the narrower, more detailed, scans. [Triesmann 88] suggests that the basic 'features' extracted by the eye's receptors all have alerting effects upon attention. She notes that these features 'pop out' of an image and are used to alert attention : 'simple, highly discriminable features can be automatically identified with or without focused attention, but that they are accurately located and co-joined only when attention is narrowed'. The following set of features are proposed as being features which are automatically distinguished, or 'pop out' of an image : luminance, colour, motion, texture, depth. One problem with the results of [Treismann 88] is that they are based on abstract images, which may not be representative of real world images.

Some validation of [Triesmann 88] results for real world images is provided by eye tracking performed by [Brandt 54], [Yarbus 67]. These used real world images and agree with the results of [Triesmann 88]. They suggest that attention will be shifted to objects that are larger in size, have greater surface detail, or are lighter or brighter than their surrounding.

Evidence of the accuracy of [Treismann 88] results can also be found within more general frameworks given for designing photographs. [Arnheim 68] describes how objects within an image can be made salient by the use of 'weight' : ' intrinsic interest has been found to be a factor of compositional weight.' He suggests that a number of factors effect 'weight' :

- Objects which are shown in more detail or of interest will be more likely to be infocus : 'An area of a painting may hold the attention of the observer either because of the subject or by its formal complexity, intricacy.'
- Larger objects will be focused on in preference to smaller
- Object which are more brightly coloured will gain focus : 'Bright colours are heavier than dark ones'
- Objects which are apart from others are focused on : 'Isolation makes for weight'
- Objects which are nearer the front of an image will be more likely to be focused upon : 'Overlapping establishes a hierarchy by creating distinction between dominating and submissive units. A scale of importance leads through intervening steps, from foreground to background.'

A similar set of guidelines are also given in [Focal 76], suggesting that the importance of an object is controlled by three factors :

- Tone : if tone is lighter or darker than surroundings then it will be in focus
- Sharpness : focus will be given to objects which are sharp and detailed
- Scale: focus is given to largest object.

One difficulty with the guidelines provided by [Focal 76], [Arnheim 68] is that they are based on rules of thumb, driven by techniques used in photography or painting. As such they offer no validation or rating of the importance of particular effects.

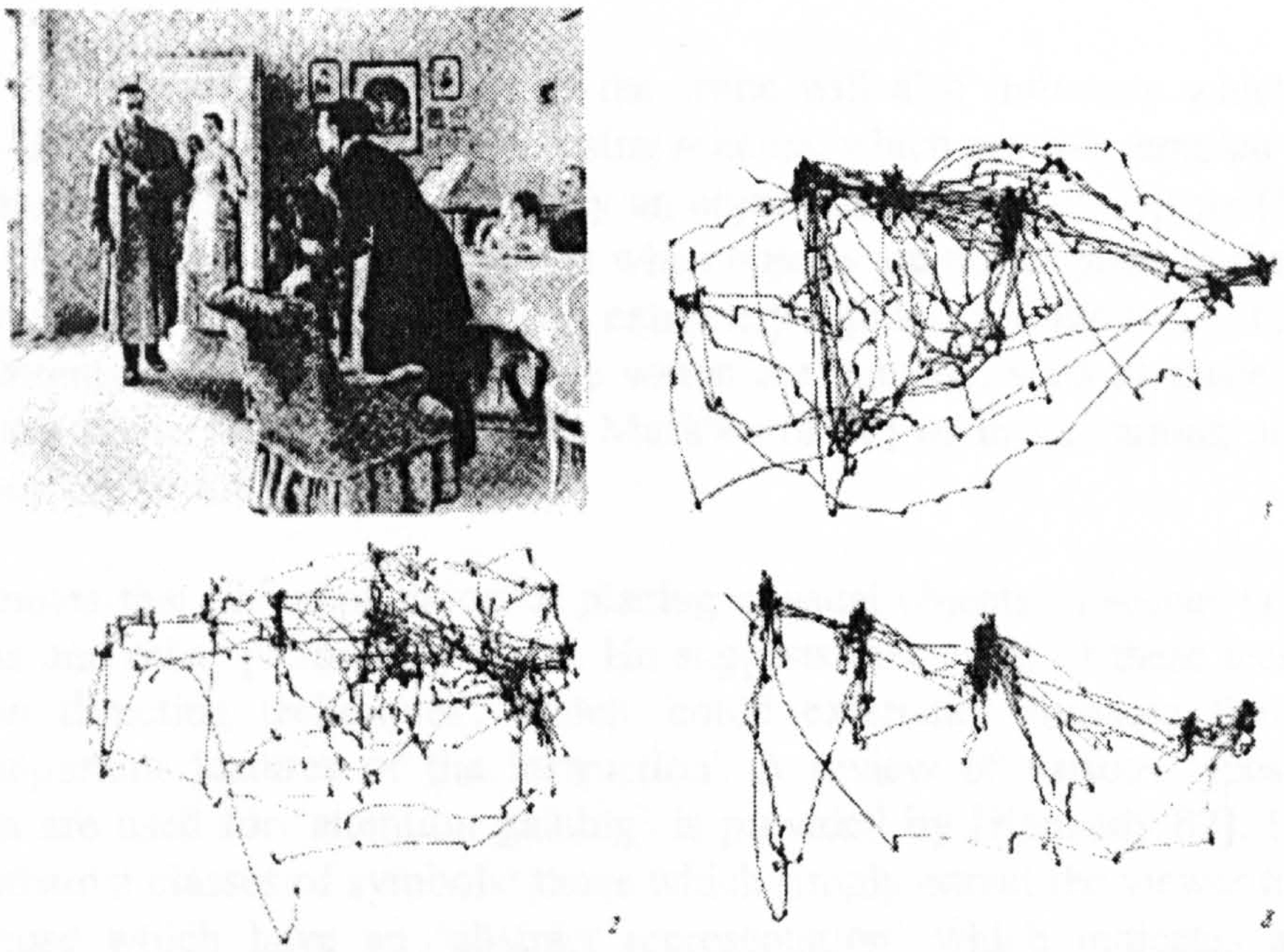


Figure 2.3.5 Eye tracking based on different tasks [Yarbus]

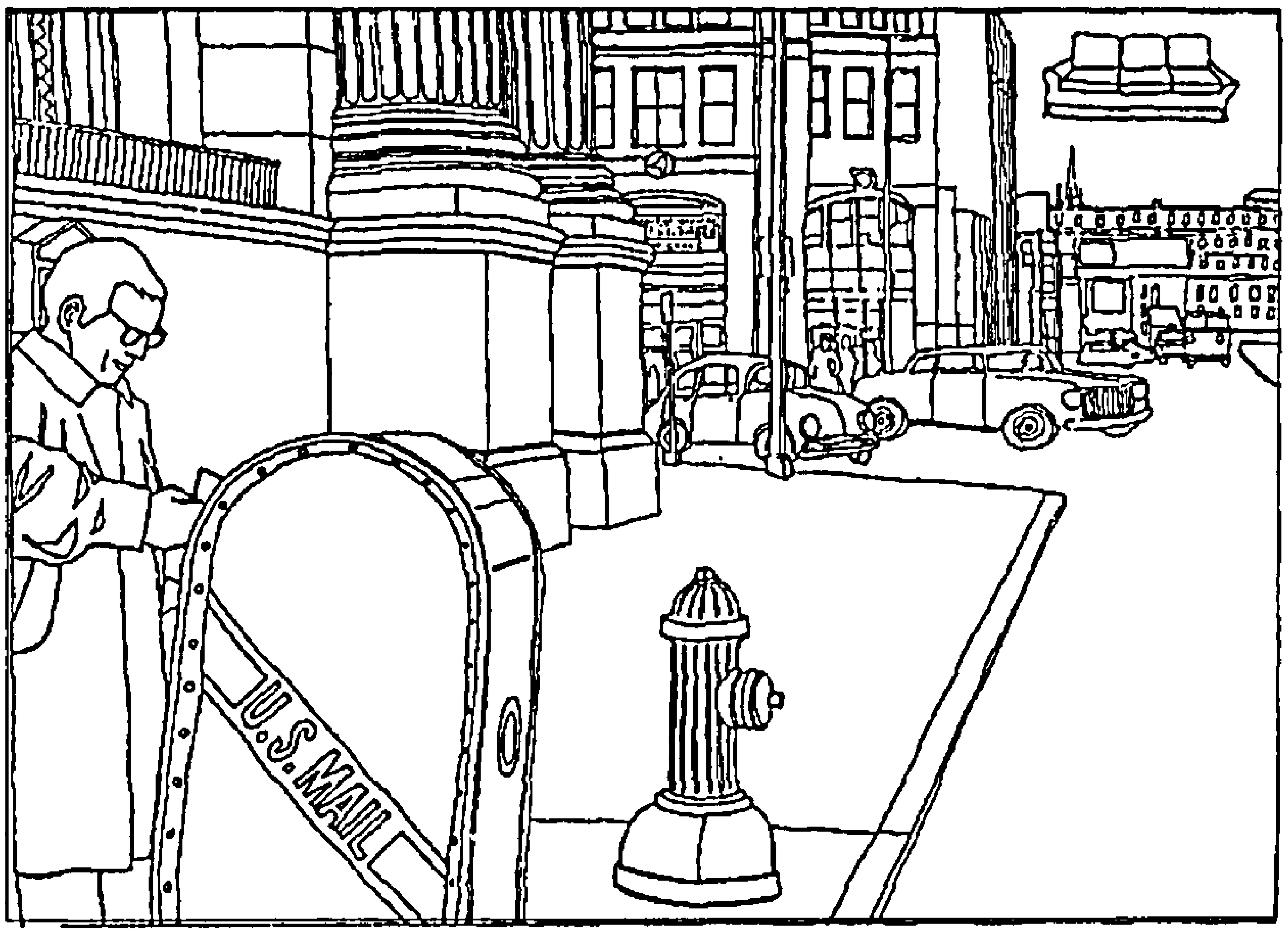


Figure 2.3.6 Unusual objects attract attention [Biederman]

Another problem with the approaches of [Focal 76] & [Arnheim 68] is that they ignore the role of the viewer's goal and expectations in dictating what will be fixated upon. The viewer's task or volition will effect which objects are fixated. [Yarbus 67] demonstrated this with a series of eye tracking studies (figure 2.3.5), in which the viewer was given different tasks, such as counting the number of people in the room.

The nature of the viewer's expectations of the scene will also influence which objects are likely to be fixated. Objects are recognised using schema, which provide templates made up of expected combinations of features to identify an object and groups of objects [Kahneman & Treisman 92]. [Biederman 81] has found that when objects are out of place (either spatially or physically) within a scene, they are noticed extremely rapidly, see figure 2.3.6. The objects processed by attention are likely to be those which are unusual, such as those being out of keeping with the scene category [Loftus & Mackworth 78] or in containing an unexpected detail or location [Pezdek 88].

[Fleming 87] notes that one application of placing unusual objects in scenes is found in the design of icons and other pictorial symbols. He suggests that value of these techniques is as 'overt attention directing techniques', which 'could externally simulate the focusing of attention to important features of the instruction'. A review of various types of pictorial symbols which are used for 'attention gaining' is provided by [Kennedy 82]. He notes that there are two distinct classes of symbols; those which simply attract the viewer to a part of an image, and those which have an 'abstract representation' which indicates an additional meaning. Thus a highlight would attract attention, whilst an arrow or pointer would have an additional meaning, directing the viewer that an important object was located at its head.

Summary

- Attention is a serial process and will take time to process an image; it will start by building up a general representation, then extract more specific details if required.

- By default only the scene level will be focused upon unless the user is driven by a particular goal; or an object in the scene is unusual or out of keeping.

- Objects which are more likely to be focused on will be those which are : bright in colour, apart from other objects, larger in size, shown in more detail, sharper or more in focus, or nearer the front of the scene.

- Unusual objects, such as pictorial symbols or icons will attract attention.

2.3.4 Animation

[Treismann 88] suggests that motion is detected within the region fixated upon by the eye by low level receptors, similar to those that extract other visual features. She suggests that motion detection does not require the use of the attentional spotlight and is possible in both the fovea and parafovea, causing the object in motion to 'pop out' of the display.

[Tepin & Dark 92] argue that the initial 'registration' of change in the visual field will be automatic, but that subsequent processing by attention for identification will not be. [Tepin & Dark 92] note if attention is already focused and in use for identification then attention may not be available immediately. It was found that in a priming task where an abrupt onset cue is given to a target location, attention will not always be oriented to the change. It is suggested that the sometimes attention is non-automatic and may also be controlled by task or volition. However, if attention is unfocused then it will be attracted automatically to change in the image.

[Hillstrom & Yantis 94] have investigated how different types of motion effect attention. They suggest that attention will be drawn to an object if motion is topographic i.e. a new object appears, or breaks from an object which was previously processed as a whole : 'when change occurs (e.g. when something begins to move) the scene may be re-segmented'. A study found that if the object in motion was different in appearance from the global scene, then recall of its identity was good; otherwise it was poor. However, the scene used was made up of letters, rather than real world objects. They conclude 'that the appearance of a new perceptual object, and little else, captures attention'; and suggest two types of events cause a new object file to be created :

(i) 'the abrupt motion onset of an object in a previously blank location';
(ii) 'the segregation of an object from its background when motion begins'.

One problem with both the results of [Tepin & Dark 92], [Hillstrom & Yantis 94] is the use of abstract materials in testing how attention would respond to motion. Unfortunately, no further studies were conducted which used real world images.

Once attention has registered a change in the image, eye tracking results from [Yarbus 67] provide evidence of a different type of scanning to that found for static images. This is termed 'pursuit', in which the eye follows the object in motion, so that attention is held on it as it moves. Pursuit may prevent attention from switching to other objects in the image as the eye tracks the motion of a particular component.

A different type of effect is achieved by changing the image completely, rather than a particular object moving within it. [Reeves 93] has studied the effect that changing from one image sequence to another in animation has on attention. These changes are commonly called 'cuts' : 'At a surface level, cuts produce a shift in orientations, edges, colour and luminance'. [Reeves 93] studied response times to visual cues before or after cues. He found that responses were more rapid following a cut. He concludes 'cuts serve as visual markers alerting the viewer to new information being presented, resulting in an orientating response of attention'. He also notes that viewer's goal has an effect, if the scene can be related by being meaning to the one prior to the cut, then less attention was found to be required. He proposes the following rules for the effects of cuts on attention :

- Attention to material presented after a cut will be greater than that presented precisely on a cut
- Attention to unrelated segments following a cut will be greater than related segments following a cut.
- Related segments require less attention and can be integrated in less time than for unrelated segments.
- Attention to related sequences will decrease over time, whereas to unrelated sequences in will remain constant.

Similiar findings are suggested by [Barnard & May 95] based on the ICS cognitive architecture given by [Barnard 85]. They propose that if a cut is made to an unrelated sequence, it causes a problem in understanding the meaning of the sequencing : 'if a cut has been unfilmic, then the blending of the propositional streams will fail', 'the viewer would have to interrupt the higher level comprehension' to work out what had happened. This suggests that cutting to an unrelated sequence will cause problems with comprehension of the sequence, leading to an 'interrupt' which will cause the viewer to increase attention to the sequence.

The studies conducted by [Hillstrom & Yantis 94] suggest that motion is only partly a basic property detected by the eye's receptors. The results of [Tepin & Dark 92] conclude that shifting attention to an object in motion may be a default if attention is not focused elsewhere, but otherwise shifting may be delayed until attention is free. They also noted that volition may control whether attention is given to motion. The results of [Reeves 93] give further details as to how attention behaves when one animation is changed for another, suggesting attention will be alerted by a cut, particularly if the material is unrelated to the previous sequence. [Barnard & May 95] add that this may interrupt comprehension.

Summary

- Attention will be drawn by default to the object in motion

- If attention is already focused it may not automatically shift to an object in motion.

- An object in motion will be tracked and may prevent focus shifting elsewhere for during the motion.

- Two forms of motion are possible within images : topographic movement of the object within a scene, or internal movement within the object of its parts. In general topographic motion, in which either an object appears or moves across the ground of an image, is more attention grabbing than internal motion.

- Cutting from one animation sequence to another will re-orientate attention and cause it to process the new scene in more detail than the scene prior to the cut, particularly if the new scene is unrelated.

2.5 Attention to Language and Image media

2.5.1 Text & Speech

Some general guidance on when to use text or speech can be found in [Deathersage 72]. He suggests the following guidelines should be applied, shown in table 2.5.1.

Use Speech	Use Text
The message is short	The message is complex
The message is simple	The message is long
The message will not be referred to later	The message will be referred to later
The visual system is overburdened	The auditory system is overburdened

Table 2.5.1 Guidelines for applying speech or text [Deathersage 72].

Whilst the guidelines provided by [Deathersage 72] are useful for making decisions between text or speech, they offer no advice concerning how to combine the media together, or what value the two media may have when combined (see 2.6.2).

[Sinitra 90] provides a summary of how reading and listening interact. She suggests that the processes of listening and reading converge after a word has been identified. In a study, she found that reaction times were better for reading when subjects first heard a matching auditory stimulus for sentences, syntactic nonsense strings and random words - but not for non-words. She concluded that reading and listening converge at the lexicon which identifies a particular words meaning.

The results of [Sinitra 90] suggest that it is important that text and speech reach the lexicon in the same form; and predict that differences in wording between speech and text may cause problems in processing. Studies by [Grimes 91] have shown that comprehension of text and speech is impaired if the two media do not convey the same information. [Grimes 91] suggests that if text is presented visually, then 'subjects can't help but to read it', 'only exact duplicates on the screen of what is being said in the narration appear to allow text to be processed unimpeded'. In summary, these studies suggest that if the lexical representation of text and speech are inconsistent, comprehension will be reduced.

[Rubin 80]; [Nickerson 80] suggest that speech and text may have different roles in a presentation. They provide reviews which compare the value of speech and text in more general terms. [Rubin] notes that text comprehension is aided by its permanence and structure :

- Readers can scan back over pages to re-read a sentence mis-parsed.
- Effective readers can also skim ahead and use section and chapter headings to find material.
- Visual organisation of text provides additional structure :
  - Paragraphs give additional information not found in speech as to the larger constituent structure;
  - Italics and underlining may be used to emphasise or contrast words.

[Rubin 80] concludes that speech may be best used where reading is difficult, or where words are unfamiliar. [Nickerson 80] provides a similar summary of the value of text over speech to [Rubin 80]. [Nickerson 80] concluded that text does not make same level of demands on WM as speech : 'the printed page constitutes a memory of sorts; when ambiguity is encountered in text it can be reread'.

Summary

- Speech is useful for short, simple information, which will not need to be re-referred to.

- Text is useful for more complex information because it can be re-read, and provides clear marking of paragraphs and titles for structure.

- If speech and text are used together, it is vital that they are tightly synchronised, and that they share the same wording

2.5.2 Text & Image

[Loftus 83] gives a summary of the main differences between text and image perceptual processing. She notes that the visual acuity required for text in a common size font is far greater than that required for an image - objects in scenes are usually large enough to be partially identified in the parafovea (possibly upto 7.5 degrees), whereas text will only be processed within 2 degrees. Fixation durations will also tend to be longer for words than for pictures; however, perceptual processing of words will be completed earlier in the fixation, whilst processing in images will continue for the duration of the fixation.

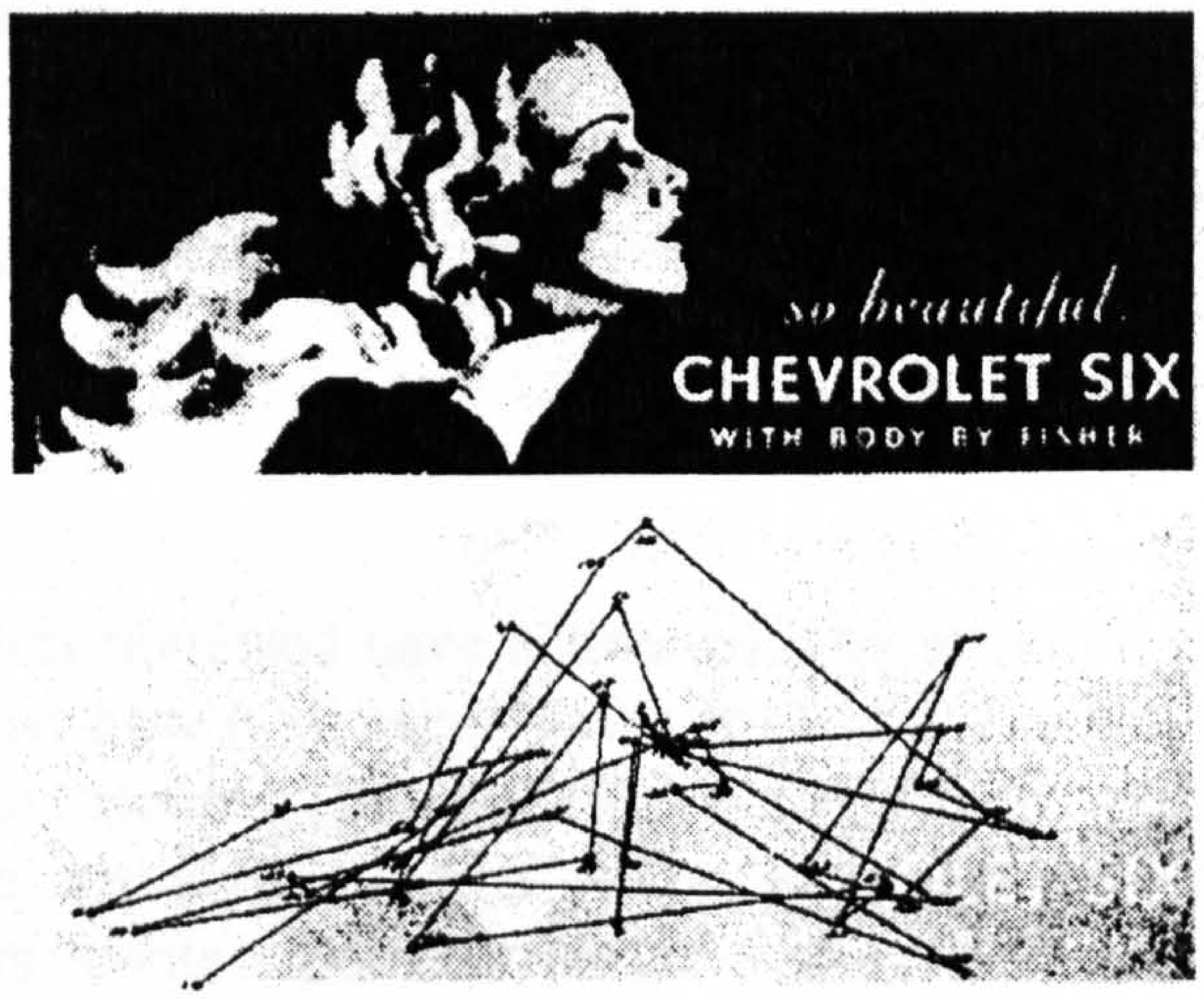


Figure 2.5.1 Eye track of text and image [Brandt]

[Theios & Arnheim 89]; [Brandt 54] propose that the order of attentional processing between text and image will to some part depend on surface size of presentation. If the text is far smaller than the image (which is usually the case for a caption or title), then the image will be more likely to be fixated upon first. [Brandt 54] provides a series of eye tracking studies concerning fixations to newspaper adverts (see figure 2.5.1). He notes since attention is in general attracted to higher detail and colour, an image will have an additional advantage in first capturing attention. [Brandt 54] concludes that 'pictures and figures have an exceedingly high attentional value; reading copy is only a last resort in many cases.'

Whilst text may not first capture attention, studies by [D'Ydellawe et al 93]; [Duffy 93] suggest that once attended to, it is very difficult for the text to be ignored. [D'Ydellawe et al 93] studied fixations given to subtitled television programmes. They found that there were no differences in conditions with or without sound in subtitle reading; though 'it would seem logical not to read the subtitle if one understands the language'. They conclude that 'reading subtitles is an automatic process', even in the case the subtitle gave no extra information.

One problem that [D'Ydellawe et al 93] note with the literature concerning subtitling is that the amount of time to read a subtitle is difficult to calculate exactly. They noted that it is normal practice in TV production to give 6 seconds to process two lines of text (64 chars) for subtitling, even if the subtitle was not informative. They suggest that the time ignores any idea of subtitle complexity, or familiarity with the language used.

In further eye tracking studies [Duffy 93] provides similar results to [D'Ydellawe et al 93]; she notes that once subjects fixated a text, they never broke off fixations to it until they had read the whole clause of text. She concludes that text has a 'lock on' effect upon attention, and proposes that this may be required by the linguistic processes which extract meaning.

Summary

- In general an image will be focused on before a text unless the text size is larger than the image area, with focus then shifting to the text after the image.

- Once attention moves to read text, it will lock on to read the a whole clause of text. This will make it difficult to focus on any changes in the image.

- Reading time for text is greater than that for image; however, only default values are available.

2.6 Contact Points

Whilst the previous studies reviewed have addressed how attention is given to one particular media, they do not address how it switched between media which are available concurrently. In a review of picture- text studies, [Baggett 89] has noted the importance of co-reference, or 'contact points' between visual and verbal information in allowing their content to be understood. She suggests 'contact points' provide a way for the image to be co-referenced with the accompanying language, so that a single representation is formed, representing the information from both media. The following sections review studies of how contact points may be formed between text and image, and speech and image.

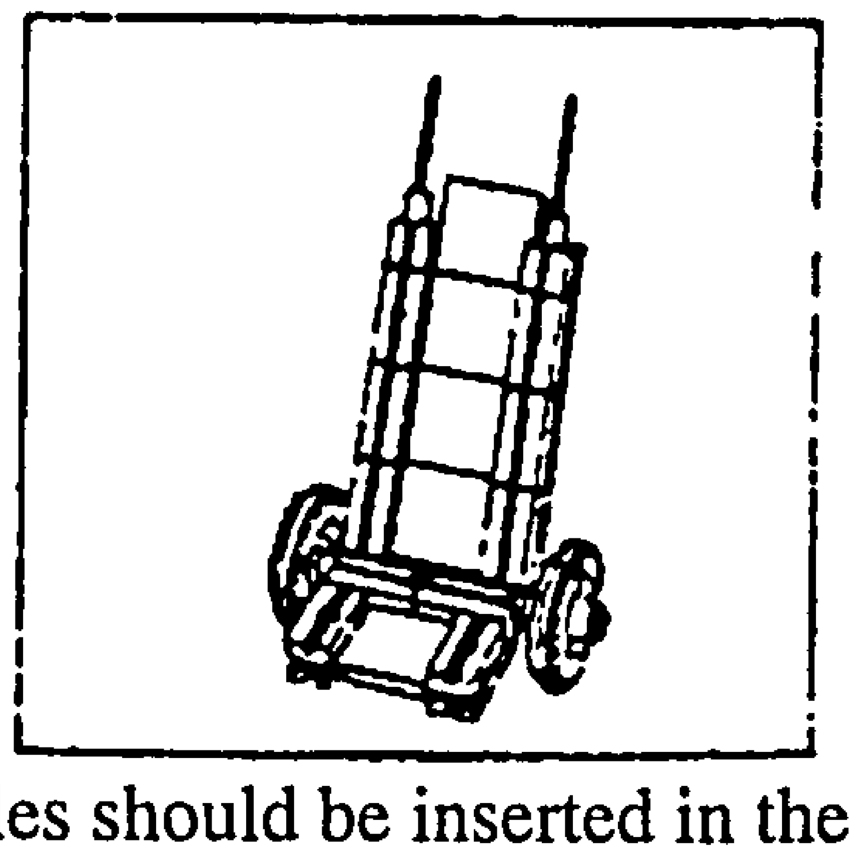
The key issues for multimedia design concern how to combine media together to form a contact point :

- How to provide linking references and synchronisation between media streams.

- How to ensure that the message thread can be followed between the two media.

2.6.1 Contact Points between text and image

Several eye tracking studies [Stone & Glock 81]; [Hegarty & Just 93] provide evidence for the formation of contact points between text and image, in which attention switches between the image and text to form or resolve co-references.



Next the handles should be inserted in the end of the back

Figure 2.6.1 Example of materials used by [Stone & Glock 81] : attention switches back and forth between the text and image

[Stone & Glock 81] performed an eye tracking study of subjects reading a set of instructions describing how to build a toy cart (see figure 2.6.1). They found that subjects first looked at the illustration for 1-2 seconds, then the text caption. They also noted that at several points in reading the text, subjects shifted back to the picture; however, the equipment that [Stone & Glock 81] used did not collect precise data on where fixations were located.

[Stone & Glock 81] conclude that 'readers who view an illustration spend the first few fixations extracting gist information. Readers then appear to refer to the text and read for several seconds. They then periodically refer to the illustration. Readers appear to be systematically comparing the semantic content of the pictures and text as demonstrated by their frequent looking from one to the other.'

A more detailed study that did use accurate eye track data is reported by [Hegarty & Just 93], who studied the order of fixations during a mechanics comprehension problem (see figure 2.6.2). Their results suggest that subjects read the text in clause chunks, then switched to the diagram. They conclude that the text clause controlled diagram inspection. They also separate local and global diagram inspections : local are detailed and used to resolve co-references, global combine and organise detailed representation. More local inspections were found inbetween clauses, whilst global were found at the start and end of text.

Based on their results, [Hegarty & Just 93] propose a model of picture / text integration. The following processing steps are suggested :

(i) read and interpret a new clause of text,
(ii) integrate with previous clauses from text,
(iii) locate referents in picture,
(iv) integrate with representation of text.

In summary, the model predicts that readers 'construct mental models of the pulley system in increments, first integrating information about components of the system to construct local representations of these components, and later combining these local representations to construct a global representation of the pulley system.' The model thus suggests that text drives the interpretation of the image, by specifying which part of the image should be viewed, and in what order.

<u>Clauses Read:</u>

<u>Reading Episode 1</u>

1 This pulley system consists of three pulleys, (1763)
2 two ropes, (814)
3 and one weight. (1746)
1 (This pulley system consists) of (three) pulleys, (666)
4 (The upper) pulley is attached to the ceiling. (1230)

<u>Diagram Inspection 1</u>

UPPER PULLEY (483)

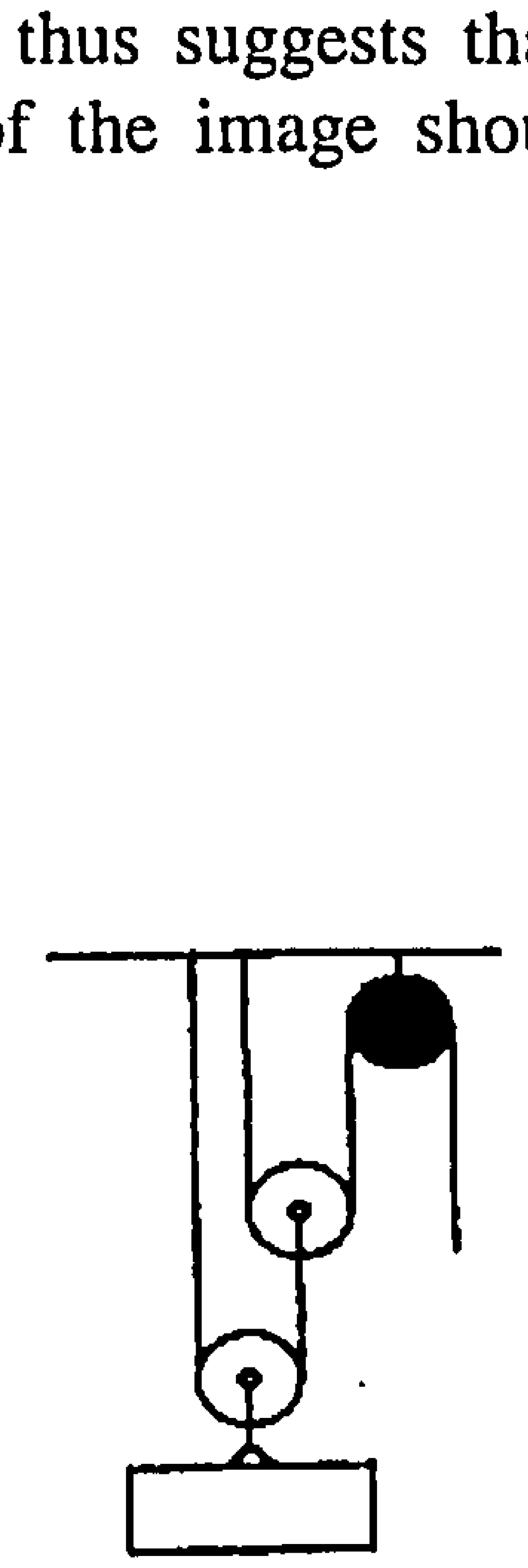


Figure 2.6.2 Example of materials use by [Hegarty & Just 93] : attention locates referents in the image given by the text.

The [Hegarty & Just 93] study suggests that their are two main issues in forming a contact point between a text and image. First the referent between the text and image needs to be located. Second the text and image must directly co-refer to the image, or else the picture and text will form two separate representations.

Several studies offer empirical evidence for the value of contact points in improving comprehension. The importance of locating the co-reference in picture-text association has also been demonstrated by [Sweller, Chandler, Tierney & Cooper 90] in the comprehension of maths problems in which parts of a formula had to be associated with parts of a diagram (see figure 2.6.3). They found that by attaching information to the relevant part of the diagram using number, arrows or labels, comprehension of the problem was improved. They conclude 'instructional material that requires learners to mentally integrate disparate sources of mutually referring information (e.g. text and diagrams) also interferes with attention and imposes a heavy cognitive load', 'when material could be combined into a unitary source of information not requiring attention splitting, performance was substantially enhanced'.They also suggest that the results are likely to be generalisable to a text with pictures.

Where co-reference between the text and picture is complex or difficult, the text may be labelled on the image, to provide a tight contact point. This will simplify the location of a co-referent. Some general empirical evidence for co-reference by labelling is given by [Smith & Watkins 72] who investigated the comprehension of descriptive passages which were either directly attached to the image to which they refer, or were separate (see figure 2.6.4). The results showed some advantage for attachment of the label within the image.

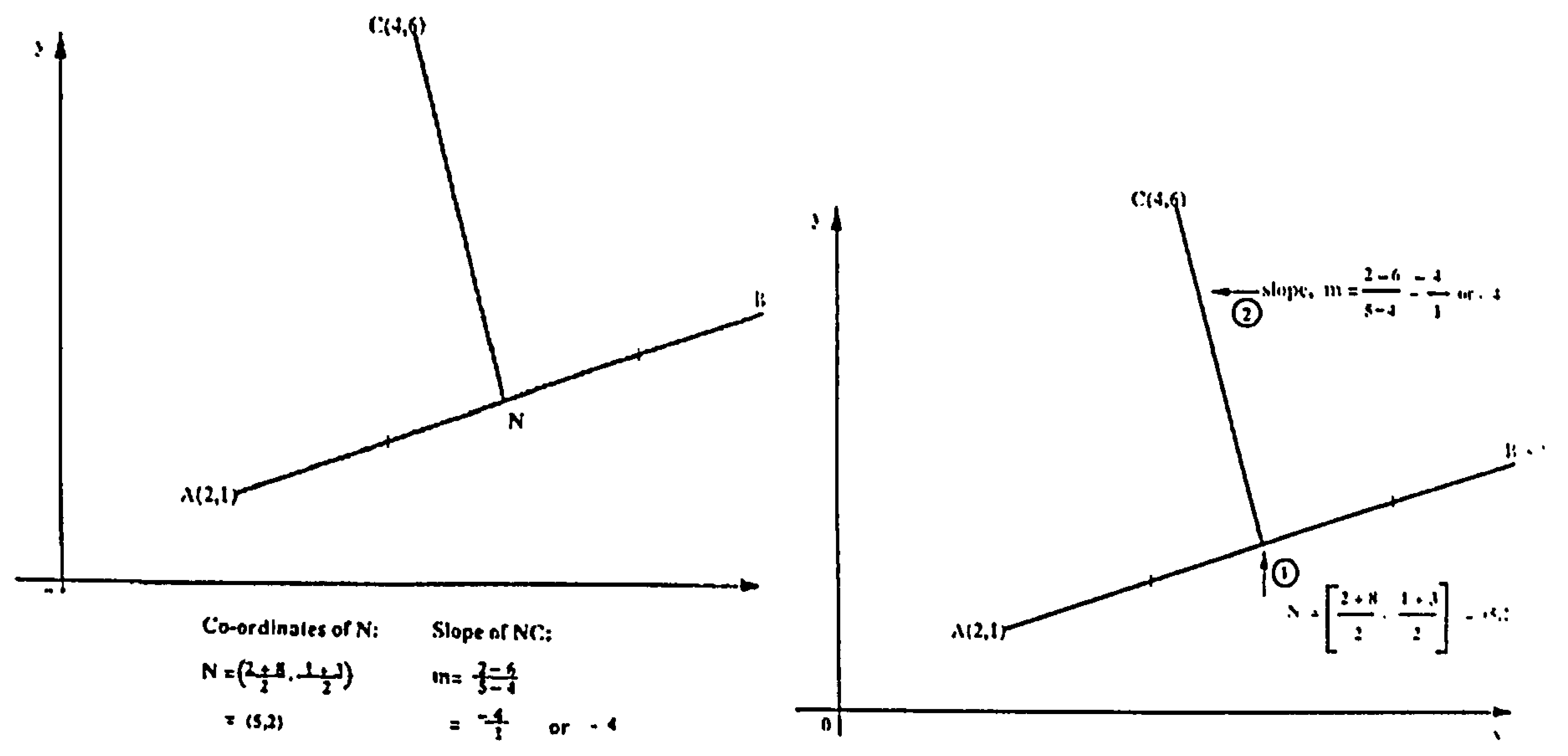


Figure 2.6.3 Example of materials used by [Sweller, Chandler, Tierney & Cooper 90] : comprehension was improved on right with labels against the left, without labels;.

These results suggest that text and image must be closely linked within the image if they are to be co-referenced. The results of [Smith & Watkins 72] suggest that this can be done by formatting the text as a caption under the image; whilst those of [Sweller, Chandler, Tierney & Cooper 90] provide evidence supporting the use of labels upon the image itself. These results suggest the importance of emphasising the part of the image which is to be referred with the text, e.g. by labelling it, using arrows or numbers.

Several studies have shown the importance of the text providing a direct reference to the image if the content of the text and image are to be integrated [Willows 79]; [Levie & Lentz 90]. [Willows 79] (cited in [Levie & Lentz 90]) investigated children's learning of information from either a picture and / or text, testing what was recalled of the content given either in the illustration alone, in the text alone, or shown in both the text and illustration. It was found that recall was improved for information given in both the text and picture than in the text or picture alone.

[Levie & Lentz 90] review a number of similar studies to that performed by [Willows 79] which compare recall of information which was given in a text and either shown in an accompanying illustration or not. They note 'with a few exceptions, the results showed that illustrations had a significant positive effect on learning illustrated text information and no effect on learning non-illustrated text material'; 'for 85% of the comparisons there was a statistically significant advantage for the illustrated text condition, and in no case was the text-alone condition better'.

It is also noted by [Levie & Lentz 90] that studies by [Pressley, Levin & Hope 81] found that learning varied by how tightly the text and image co-refered : 'learning varied with the degree of picture-sentence agreement'; 'illustrations with a high correspondence to the verbal message result in higher retention than those with lower correspondence'. [Levie & Lentz 90] conclude 'pictures facilitate the learning of accompanying verbal information that is highly related to the information pictured'.

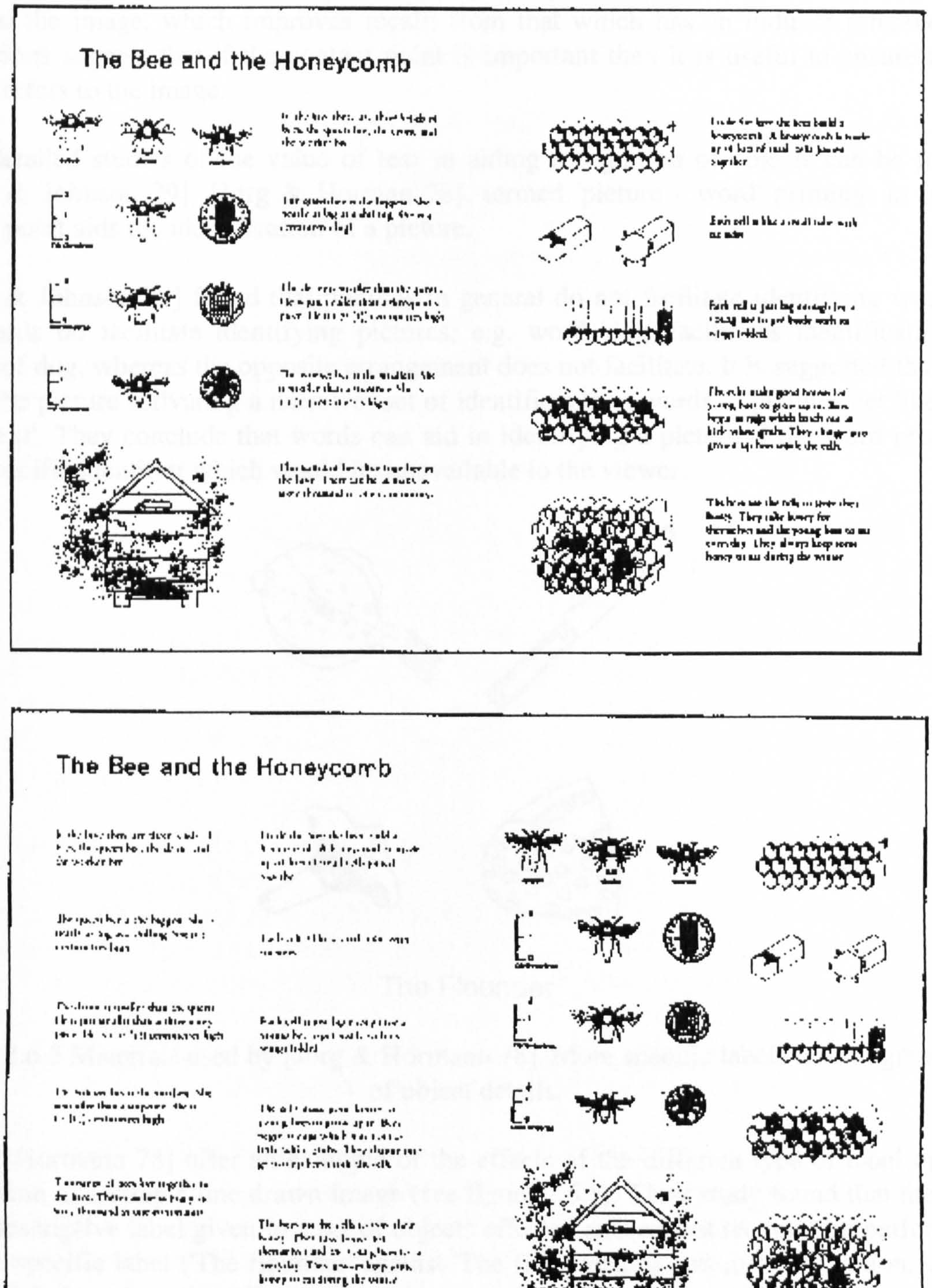


Figure 2.6.4 Example of materials used by [Smith & Watkins 72] : (top) text referencing image; (bottom) text separated from image. Referencing text and image close together aided comprehension.

The results of [Levie & Lentz 90]; [Willows 79] clearly differentiate between text that directly refers to the image, which improves recall; from that which has an indirect reference. The conclusions suggest that if the contact point is important then it is useful to ensure the text closely refers to the image.

More detailed studies of the value of text in aiding recognition of objects can be found in [Durso & Johnson 79]; [Jorg & Horman 78], termed 'picture - word' priming, in which a contact point aids the identification of a picture.

[Durso & Johnson 79] found that pictures in general do not facilitate identifying words, but that words do facilitate identifying pictures; e.g. word 'dog' facilitates identification of a picture of dog, whereas the opposite arrangement does not facilitate. It is suggested that this is due to the picture activating a narrower set of identifiers than words - 'pictures act like words in context'. They conclude that words can aid in identifying a picture if the word provides a more specific identifier which would be unavailable to the viewer.

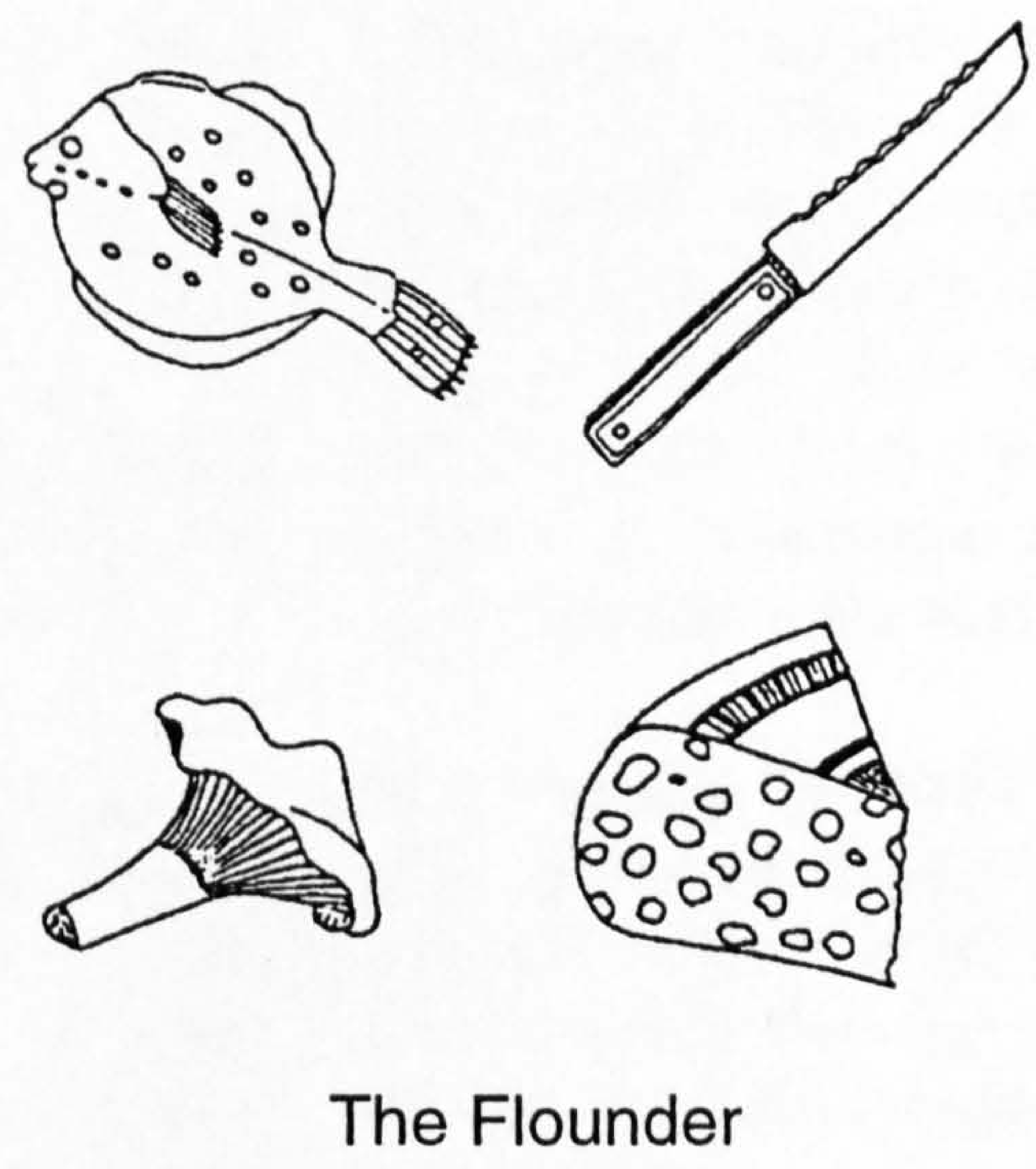


Figure 2.6.5 Materials used by [Jorg & Hormann 78]. More specific labels lead to great recall of object details.

[Jorg & Hormann 78] offer an example of the effects of the different type of label upon the recognition of a simple line drawn image (see figure 2.6.5). Their study found that the nature of the descriptive label given to a set of objects effected subsequent recognition performance. A more specific label ('The flounder' against 'The fish') lead to less mis-recognition, whilst a simpler label made little difference to mis-recognition against an unlabelled picture Further analysis also found that the recognition of other unlabelled objects was also improved by the use of the more specific label.

The results of [Durso & Johnson]; [Jorg & Hormann] provide evidence that a text contact point is useful in addition to image for object identification, particularly if the image is to be identified as a specific type of object, such as in the [Jorg & Horman 78] results.

Summary

- Combination of text and image should have clear points of contact, or cross referencing.

- If the connection between information in an image and language is important ensure that the text and image are tightly linked by shared content. This may be achieved by using a 'direct' reference, referring to the objects identifier or name.

- It is important that referents can be located. Emphasising parts of the image helps the viewer locate the contact point.

- A contact point is useful for object identification, particularly if the image is to be identified as a specific type of object

2.6.2 Contact Points between speech and image

Several studies have been conducted on language combined with visuals. [Grimes 91] has found that 'auditory-visual dissonance' may cause problems in comprehension. A study was conducted which showed a video of a news story, together with a speech track which was either matching or different to the video. It was found that the auditory and visual information were comprehended together if they 'belonged' together; if the auditory and visual information were different then auditory memory was poor, but visual memory was good. [Grimes 91] concluded that 'if two channels are perceived as a semantic unit, they should require less attention because it does not have to be distributed between them both'

In a subsequent study, [Grimes 91] used a probe in the audio or visual channel to test the attentional load during presentations with either matching or different audio-visuals. The probe had to be responded to when shown, and reaction times to it were analysed. He found that the visual probe was responded to more rapidly than the auditory probe : 'Apparently if unrelated visual and auditory stimulus are competing for simultaneous processing, then visual encoding will have primacy'. If the two channels were correspondent then subjects paid more attention to narration, and were also able to process the visual information better. In this condition, [Grimes 91] concludes 'the narration drives the processing of visuals'.

Using unrelated audio and visual information may also cause difficulties in comprehension; [Pezdek & Stevens 84] found that recognition memory for audio material was significantly reduced by the presence of conflicting video. They note that recall of auditory material dropped to near chance when accompanied by conflicting video, but that recall of the video material remained high. They also noted that recall performance was improved when matched auditory and visual information were given, as against unmatched or separate conditions.

A more detailed study of how matched speech and picture processing may improve comprehension is given by [Wagenaar 81]. This experiment compared pictures and spoken words presented singularly, contiguously, and from same or different category. Facilitation was observed only when picture and speech were from same category and presented contiguously. He concludes that 'bisensory presentation does allow facilitation, but only under conditions in which simultaneous presentation allows a link between two distinct but relatable stimuli to be created. Presenting information in two modalities is not enough unless it is possible to create a new representation, which combines material presented separately. It appears to be important that the juxtaposition of pictures and words leads to an elaborated whole which is more than the sum of its parts'.

[Baggett 84] has noted the importance of synchronisation between visual and speech media. A study was conducted in which a procedural assembly task was shown in animation, accompanied with text which was either presented before, at the same time, or after the animation. She found that performance in the task fell if speech was given prior to the animation, but was similar if given at the same time or after. [Baggett 84] concludes 'If one's objective is associative recall of names of objects, one should present the visual part early or simultaneously with the verbal. It is better for the visual component to precede the verbal media rather than vice-versa'.

Summary

- If verbal media are being presented concurrently with visual media, the message on two media should be integrated. It is important the language and image share the same topic.

- When visual and verbal modalities are used concurrently, good synchronisation is required. The verbal portion of the message should be given at the same time or after the visual, never before.

2.7 Comprehension

Design for comprehension concerns how to select particular media, or combinations of media to effectively deliver an information type. The key issues here are :

- How to choose particular information types to represent the content of the presentation

- How to select or combine media to best present the desired information types

- How to design within the media selected to support the information types

In order to consider how different media, and combinations of media, are suited to presenting different types of information, an ontology of information types is required.

[Miller & Johnson-Laird 76] provide a detailed account of how language and vision are linked to meaning. They motivate their analysis of the types of concepts required by comparing how language and vision can be used to describe the world : 'what we have called a label is essentially a rule involving a description of objects or events that can be used to determine whether an object or event is a member of the concept to which the label applies'. [Miller & Johnson-Laird 76] suggest that conceptual categories are made up of objects, space, action, sequence and causation :

[Jackendoff 83] provides a similar ontology to [Miller & Johnson-Laird 76]. His notion of 'conceptual structures' is made up of a set of types which he suggests encompass 'a single level of mental representation, or conceptual structure, at which linguistic, sensory and motor information are compatible.' He provides structure for objects, places, actions, time and causation.

For representing task based presentations, the framework used in this thesis refines [Miller & Johnson-Laird 76]; [Jackendoff 83] categories into a set of information types. Three high level groupings are made to represent the structure of the task into descriptive, operational, and organising information as follows :

- Descriptive information concerns the appearance and arrangement of objects. It is made up of Physical, Composition and Spatial information types. Composition and Physical are introduced to capture the difference between identity as Physical information and whole-part as Composition information. Spatial is used to capture location.

- Operational information concerns actions. It is made up of Physical Action and Role. Physical Action information concerns the change or motion which takes place upon the object. Role is introduced to account for the way in which one or more objects may interact in an action within a task.

- Organising information concerns how the task is organised together. It is made up of Procedural and Causal information types. Procedural was introduced to account for how actions are observed in a task sequence, rather than a sequence in time. Causal is the high level, cause-effect relations explaining how and why some part of the world behaves.

Each of the following sections considers which media should be selected or which media should be combined to present a particular type of information, and how media design techniques can be used to support the information type required.

2.7.1 Descriptive information

Descriptive information concerns the identity and location of objects. It is made up of physical, compositional, and spatial information types. Physical and composition information are related to how still objects are identified, and how they are composed together. Physical information concerns the identity and appearance of an object. Composition is related to the organisation of parts and objects into groups. Spatial information concerns how an object is located.

2.7.1.1 Media Selection & Combination : Physical & Composition Information

The recall of physical and composition information have been investigated by several studies [Boohrer 75]; [Dwyer 67]; [Bieger & Glock 84].

[Boohrer 75] conducted a series of studies examining the performance of an identification task. Subjects had to search a control panel and find particular instruments. They were given instructions as either text, or image, or as a combination. The study found that performance was poor with text only, and was best with conditions giving a picture and text combination :

'The highly pictorial multiple channel formats consistently produced the fastest time and lowest errors'. He notes that 'the pictorial channel is better for presenting static objects than the print channel.' However, using a picture alone was found to increase errors, over a combination of picture and text. [Boohrer] concludes that text and image may aid in identifying a particular object, and may also be useful in giving abstract details about the object, such as number or state.

[Dwyer 67] conducted studies of the comprehension of a two thousand word instructional text describing the function of the human heart. The studies compared a control group which used text only materials, with several other groups using different types of visuals. [Dwyer 67] concludes that the task is important when considering the effectiveness of visuals. It was found that images were effective in providing physical detail and composition information. For identification tests, a combination of picture and text improved performance over text or picture alone. [Dwyer 67] notes that identification requires the use of both location and appearance, together with more abstract verbal information providing the name or identifier of the object.

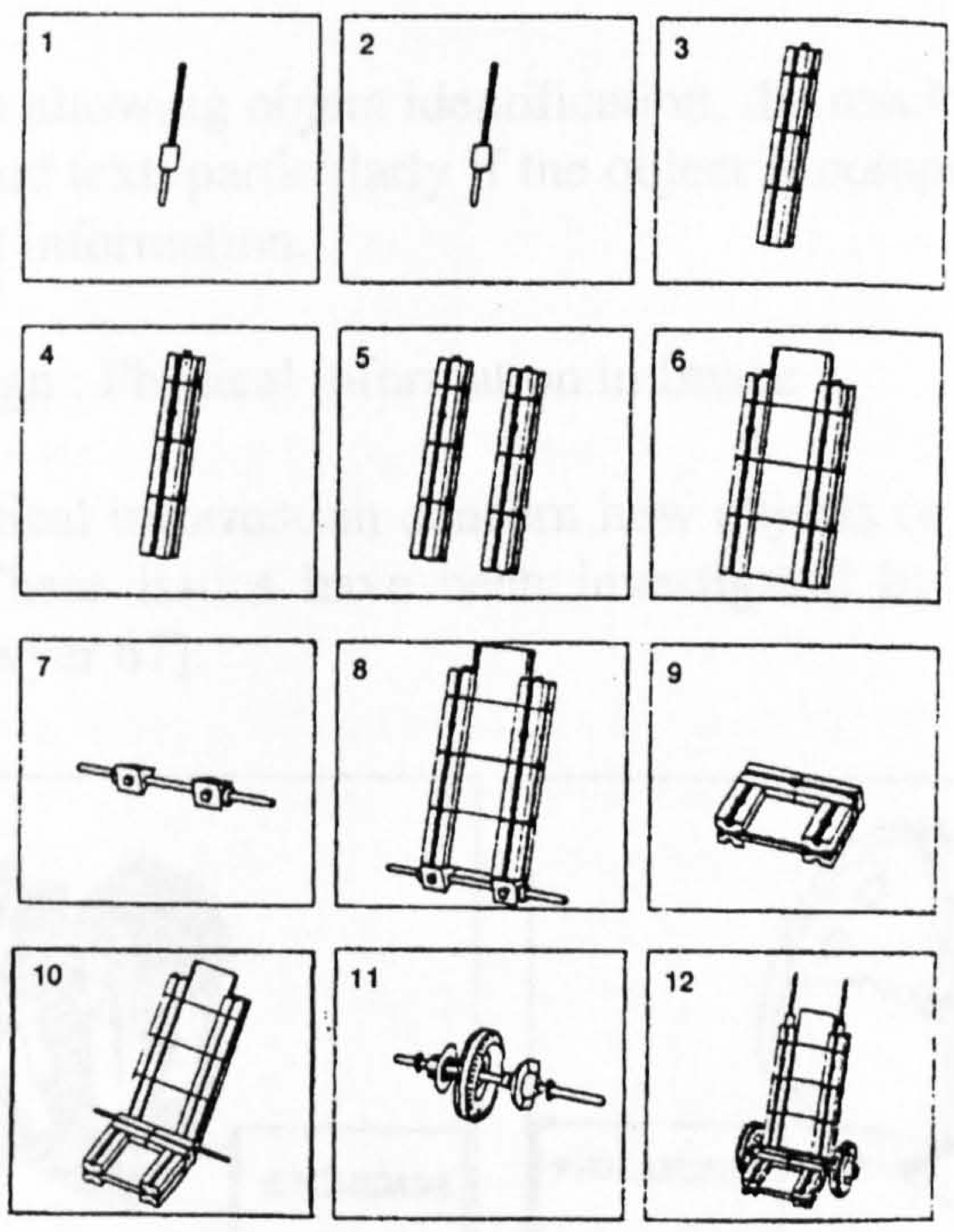


Figure 2.7.1 Picture materials used to test picture against text for an assembly task [Beiger & Glock 84]

[Bieger & Glock 84] compared performance of an assembly task with text only versus pictorial instructions for different types of task (see figure 2.7.1). They found that for identification and composition tasks, performance was increased by the use of pictures : 'there was a substantial reduction in assembly time and a slight reduction in number of errors when the contextual information was presented in pictures rather than in text.' Furthermore, it was noted that text also tended to reduce the number of errors when used to identify objects.

[Haring & Fry 79]; [Wadill & McDaniel 92] conducted studies on the how a picture effected children's recall of composition and physical detail information from a text. [Haring & Fry 79] used two sets of pictures, which showed either simple outlines, or more detailed drawings, together with a text which describing the composition of the scenes shown and also specifying particular details. They report that an image improved recall of composition information, such as which object went with which, but could find no significant difference for recall of object detail.

A further study by [Wadill & McDaniel 92] using more complex visual materials found that pictures did improve recall for object detail. They conclude 'there are boundaries to [Haring & Fry 79] original conclusion that illustrations are generally ineffective in aiding the recall of details. The present work indicates that the recall of details can be enhanced by illustrations.' These results suggest that recall of detail can be aided, but only if the details are complex and difficult to express fully in language.

Summary

- For composition information, images offer a benefit in performance, allowing objects to be grouped and placed within a context of the scene.

- For physical information allowing object identification, the results seem to favour the use of a combination of image and text, particularly if the object is complex or unusual. Text should be used for abstract object information.

2.7.1.2 Within media design : Physical information in Image

Design decisions for physical information concern how objects or scenes should be shown to improve identification. These issues have been investigated by [Price & Humphreys 89]; [Biederman & Ju 88]; [Dwyer 67].

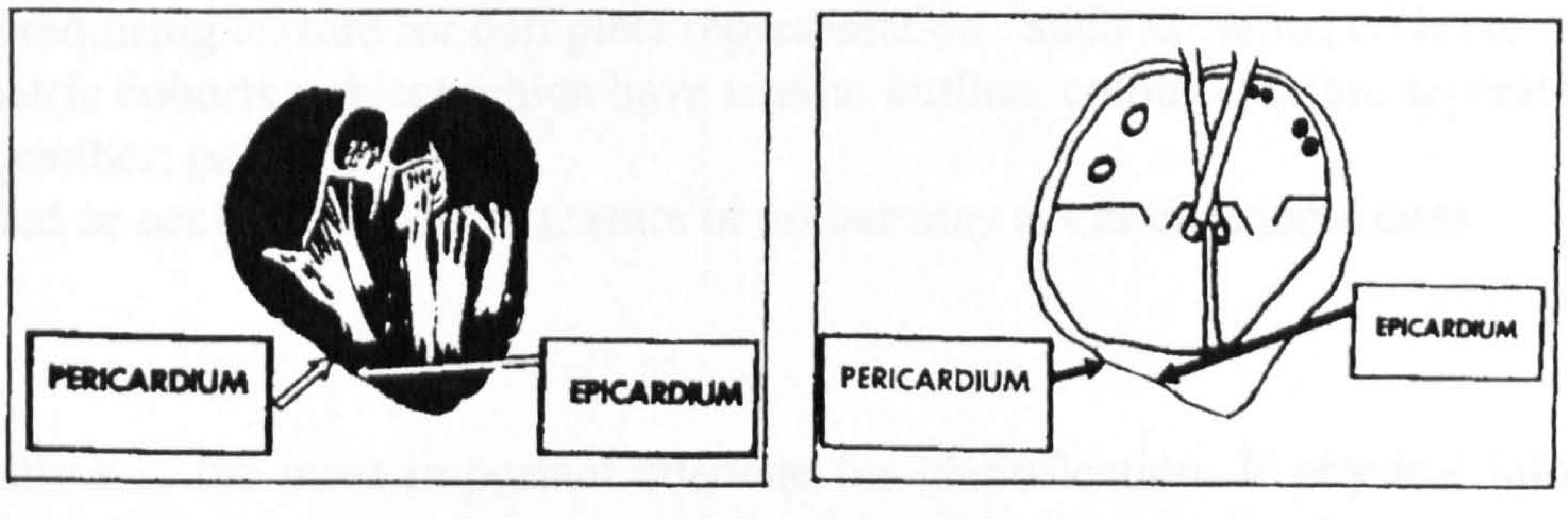


Figure 2.7.2 Materials used by [Dwyer 67]: recall was higher with simple line drawn images (right) than with colour and shading (left)

[Dwyer 67] conducted a large number of studies on the use of visuals to accompany materials describing the function of the human heart (see figure 2.7.2). [Dwyer] found that varying the type of visual image effected performance in subsequent identification and recall tasks. The following main conclusions are proposed : i) increasing the amount of the realistic detail will not produce an automatic increase in the amount of information assimilated from it; ii) aesthetically pleasing visuals may be deceptive in their instructional value; iii) the type of visual illustration that is most effective depends on the type of information to be transmitted.

More detailed results of the value of outline 'v' detail for identification are provided by [Price & Humphreys 89]; [Biederman & Ju 88]. [Price & Humphreys 89] suggest that an object will first be identified by making use of the shape and outline (see figure 2.7.3). By comparing recognition time for outline of objects against colour and shape they found that surface detail gave no advantage for categorisation; but that individualisation did require both shape and surface detail. They conclude 'knowledge of shape may give key to unlock information concerning item specific colour knowledge - stored knowledge of object colour only tapped by shape'.

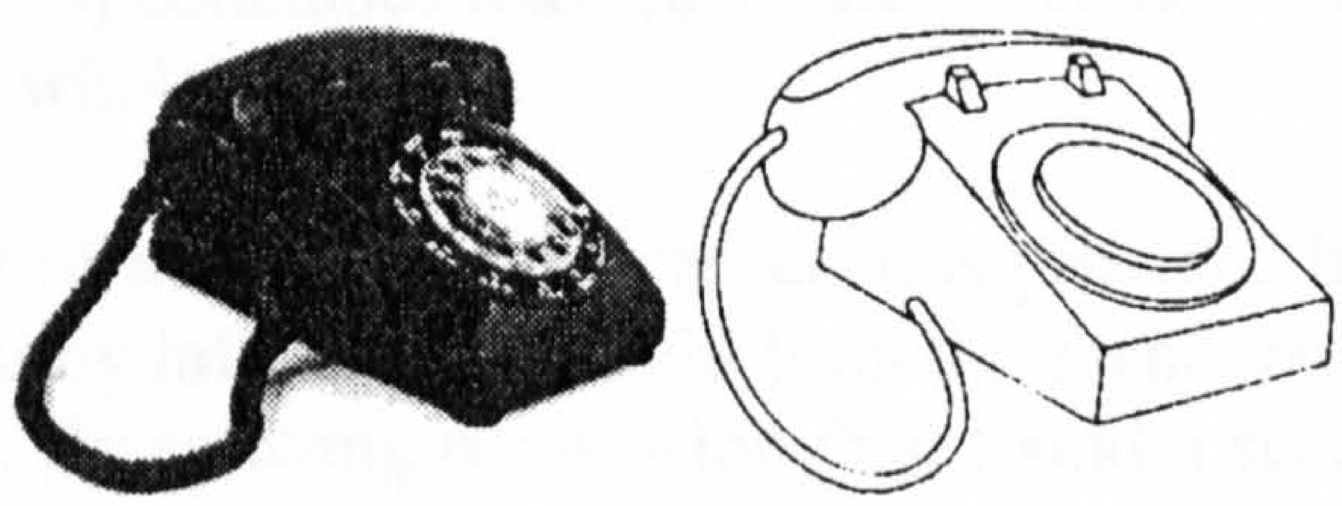


Figure 2.7.3 Material used by [Price & Humphreys 89] : recognition as rapid for an outline object (right) than as for colour and shading (left), showing that recognition is based on outline.

Similar findings are also given by [Biederman & Ju 88] who studied recognition of whole or partially occluded objects which were shown either in outline or using texture and colour. They conclude that 'although differences in surface characteristics such as colour, brightness and texture can provide cues for visual search, they play only a secondary role in the real time recognition of intact objects when its edges can be readily extracted'. [Biederman & Ju 88] also note that colour and texture may be important in certain circumstances :

(i) Mass objects : e.g. snow, water - which are not easily separable into discrete objects
(ii) Objects requiring texture for complete representation : such as screw, coils etc
(iii) Volumetric cohorts : object which have similar outline, colour / texture separates e.g. leopard 'v' panther; peach 'v' plum
(iv) Degraded or occluded objects : texture or colour may act as additional cues

Summary

- Object outline is the most important attribute for identification. If physical information is required then it is important to show the whole of the object, unobscured.

- A range of circumstances exist in which colour and texture may also be useful : if specific physical information is required, concerning a particular instance of an object to be identified, or details of object parts, or needs similar objects to be distinguished, then colour and texture should be applied.

    - Caveat : If only general physical information is needed, such as for a class of objects, then outline may be sufficient.

2.7.1.3 Within media design : Composition information in Images

Decisions for composition information concern how objects within scenes can be designed to aid recall; and how presentation techniques such as framing and zoom can aid in structuring composition.

Studies by [Pezdek 88];[Friedman 79]; [Mandler & Ritchey 77] have all found in that in recall of scenes, little surface information is retained. The general experimental method used by these authors is to give subjects an image to view, followed by a recognition test of the original image against degraded or changed distracters (see figure 2.7.5). Results suggest that the encoding of composition formed in memory is a more general representation than the original image. [Pezdek 88] concludes that scenes are processed 'only to the degree necessary to process the scene as a whole'.

A more detailed account of whole / part picture recall is provided by [Mandler & Ritchey 77] who conclude that inventory information is held from the scene, but that specific object based descriptive detail is lost. By studying recognition they found a scene may produce two kinds of composition encoding :

- 'inventory detail' of the object themselves within the scene, which was found to be adequately recognised,
- 'descriptive detail' of particular objects within the scene, which was poorly recognised.

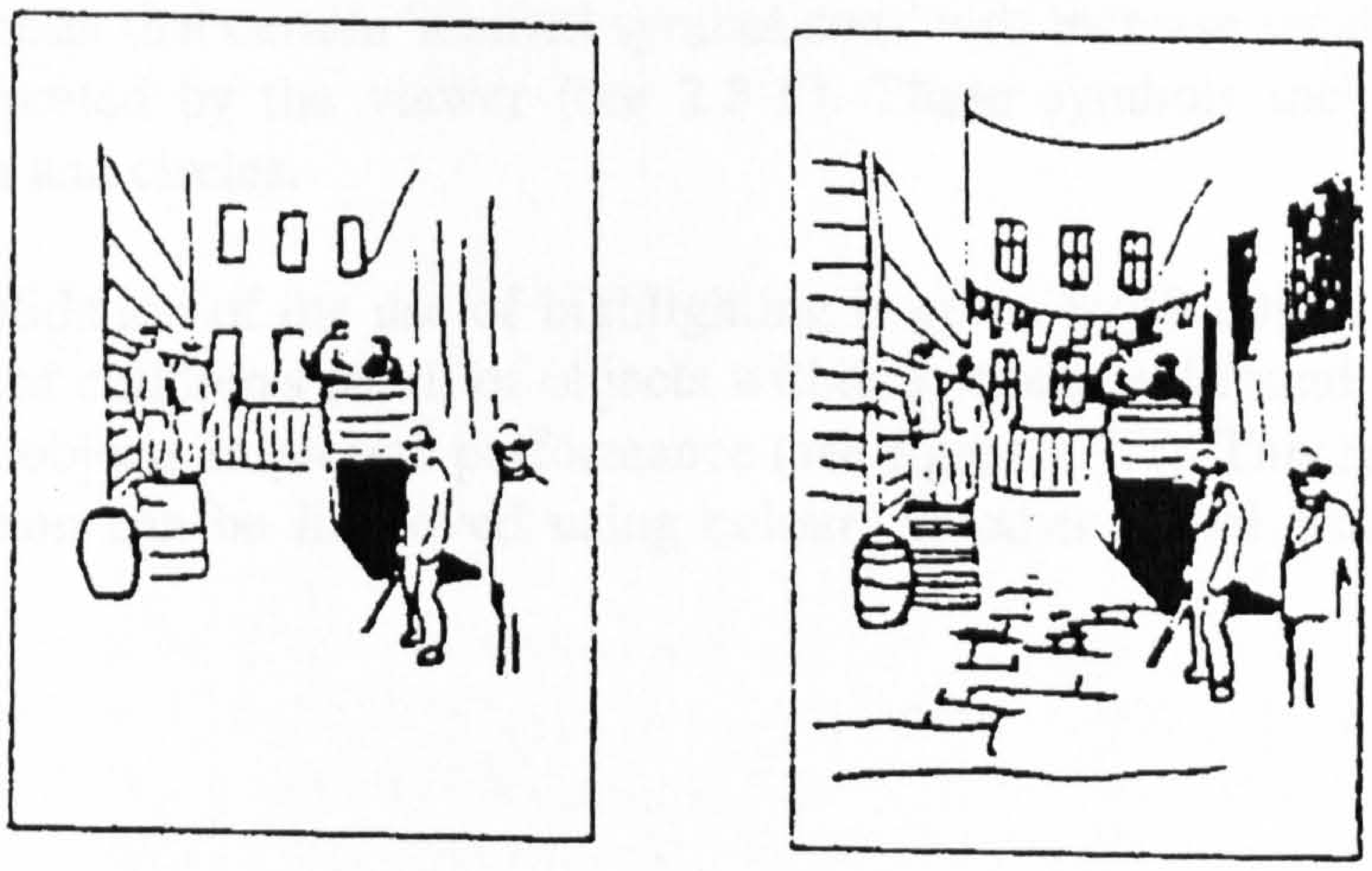


Figure 2.7.5 Materials used by [Pezdek 88] : recall of object details was poor, subjects would have difficulty remembering which objects in the image on the right had been changed or removed on the left.

Further 'whole / part' object and part detail relations will be recalled if an object is unusual or out of place within a scene, or if an object is further processed by the viewer due to volition. Many studies have noted that if an object is out of place in a scene, such as being out of keeping with the scene category [Loftus & Mackworth 78]; [Mandler & Parker 76] (see figure 2.7.6), out of expected location [Friedman 79], or an unusual detail [Pezdek 88] then it would be well recalled. [Mandler & Parker 76] concludes that 'unless changes in descriptive information violate real-world knowledge, the accuracy of their recognition may depend primarily on figurative detail'.

By using design techniques, similar results to [Mandler & Parker 76] have also been produced by [Smith & Watkin 72]; [Fleming 87]. This can be achieved in a number of ways. Learnt symbols, which are unusual in the context of the scene, such as arrows or highlights can be used to pick out an object. Alternatively, the composition of objects within the scene can be changed to make certain objects be more likely to be recalled.

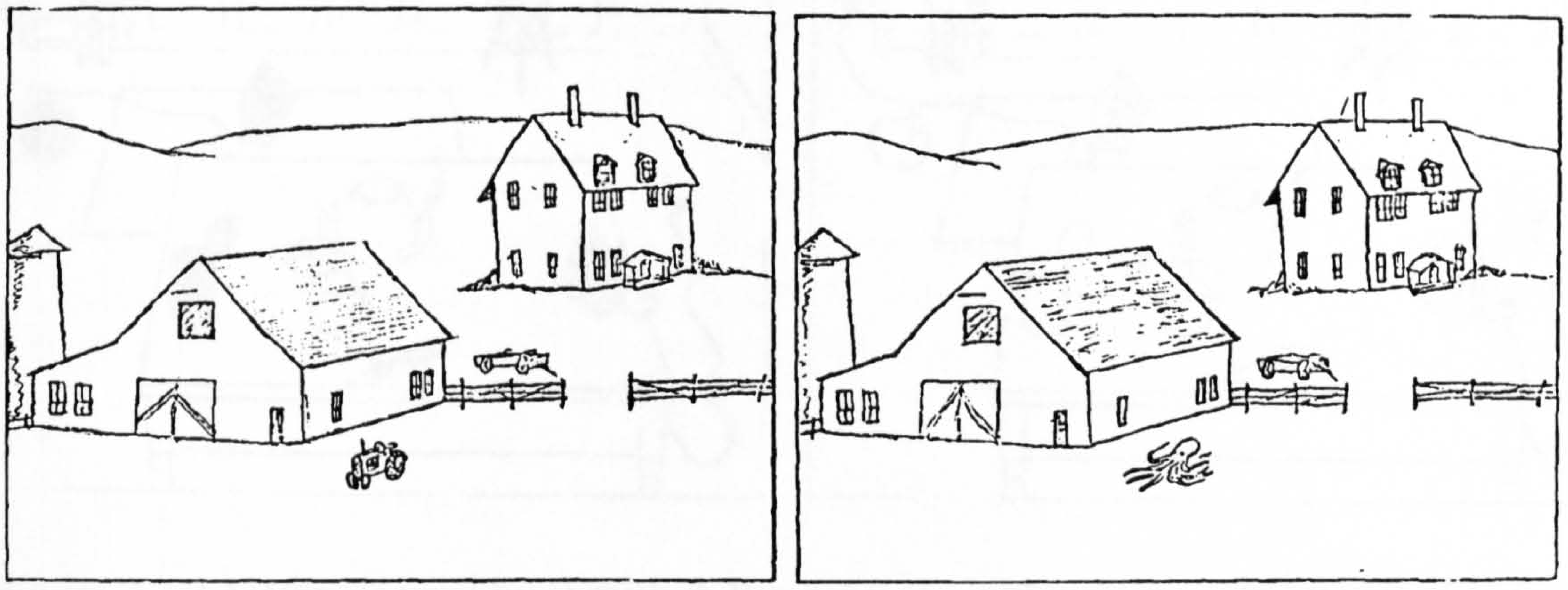


Figure 2.7.6 Materials used by [Mandler & Parker 76] : (left) the tractor is in-keeping with the scene, (right) the octopus is an unusual object in this context

[Fleming 87] suggests that certain 'learned symbol cues' will increase recall of objects if they are correctly interpreted by the viewer (see 2.3.13. These symbols include icons, arrows, highlighting, boxes and circles.

Some empirical validation of the use of highlighting is given by [Smith & Watkin 72]. They performed a study of children's recall of objects within a scene, and found that using a strong colour to highlight objects improved performance (see figure 2.7.7). This results suggests that recall of composition can be improved using colour, or other visual features which attract attention.

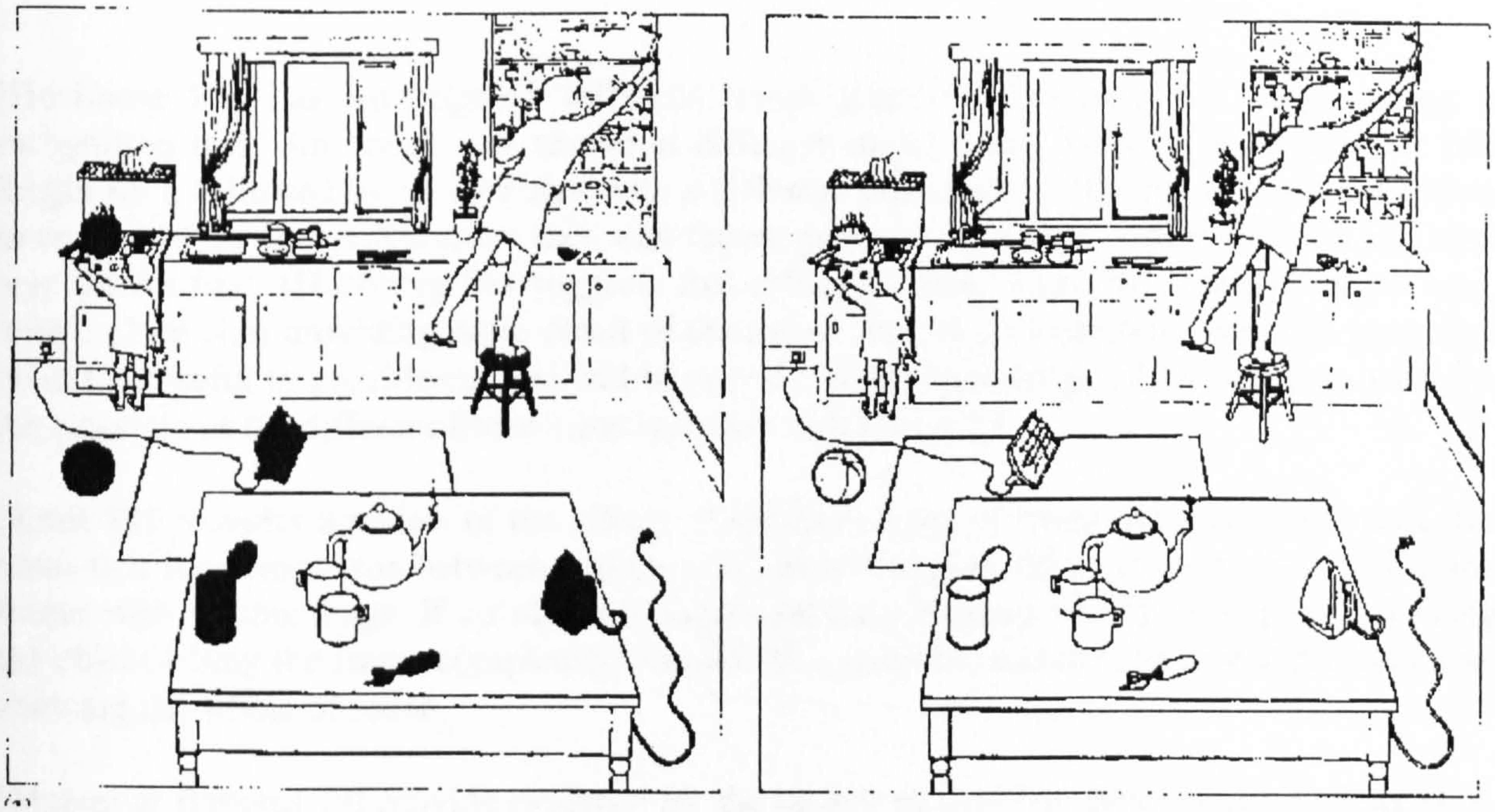


Figure 2.7.7 Materials used by [Smith & Watkin 72] : highlighting aided recall of objects

An additional method for selecting particular objects, or object groups from their surroundings is provided by the use of filmic techniques such as framing and zoom. The use of these techniques removes part of the image by increasing the size of a particular part of the image. This 'zoom' effect has been investigated by [Salomon 79]; [Hochberg 86]; [Marcel & Barnard 79].

[Salomon 79] performed a study which investigated the effect upon children's recall of using different types of pan and zoom techniques. A film was made showing a complex painting. The zoom techniques were either shown in full, giving the entire zoom; partial, showing only the initial and final state; or minimal, showing only the initial state i.e. the painting in full. The techniques were used to draw attention to eighty details within the painting. The study found that full zoom techniques were superior to partial; and that minimal was only useful in higher aptitude children. [Salomon 79] also noted that minimal techniques were only useful in higher aptitude children. [Salomon 79] summarises : 'It was not the details of Breughel's paintings nor the structure of the solid objects which was to be learned from these slides, but rather the schematic operation of singling out details or laying out solid objects.'

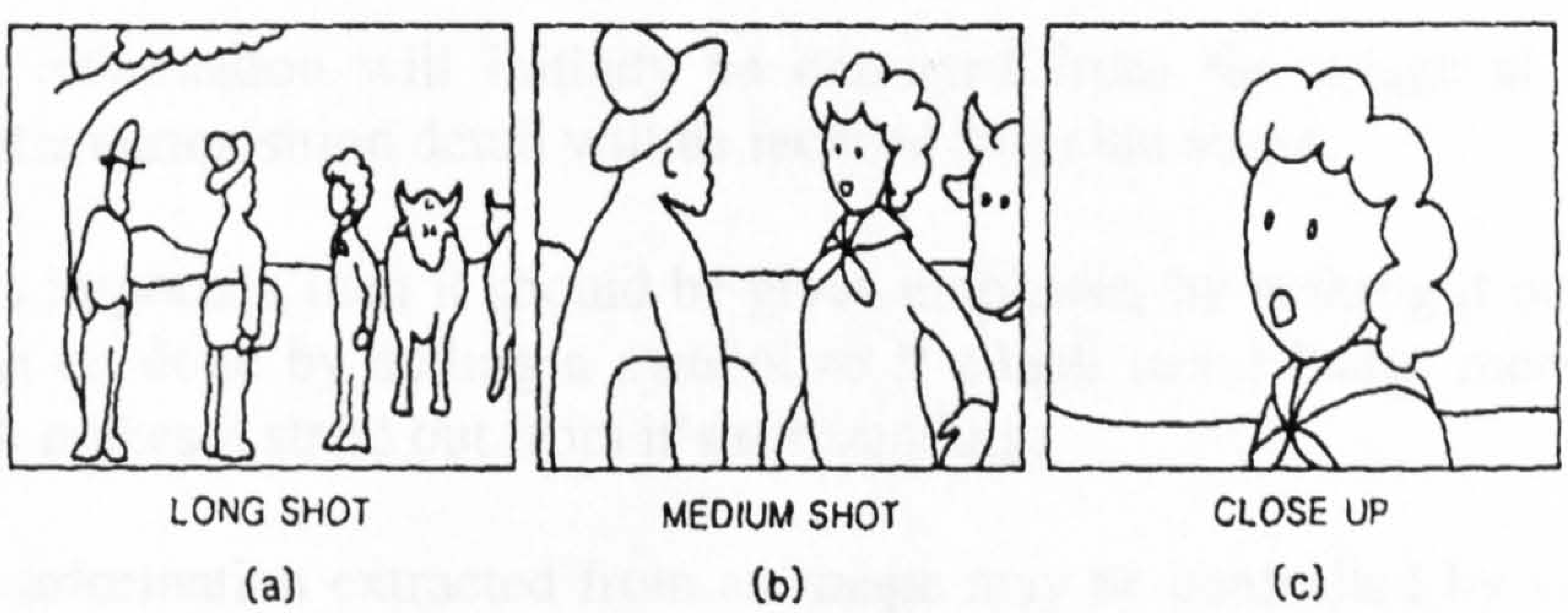


Figure 2.7.8 Example of shot types from [Hochberg 86]

[Hochberg 86] has investigated different frame sizes for showing an image, using a recognition task. An image was shown at different frame sizes, either as a close-up or full length shot, followed by another shot with a different frame size. Subjects required more time to recognise the image if a close shot was shown prior to a full shot, rather than if a full shot was shown first. [Hochberg 86] suggests that different frame sizes will have different uses, with a close shot providing more detail of the object and its surroundings, whilst a long shot would be useful to provide context and to provide initial knowledge of the scene as a whole. An example of the different frame sizes is shown in figure 2.7.8.

[Kraft 87] provides a review of the effects of different types of frame sizes and their uses. He notes that for comparison between objects a full shot is useful, showing the objects full in the frame with surroundings. If an object is important then closeup should be applied, showing the object filling the frame completely; whereas if context is required use an establishing shot showing the whole of scene.

[Marcel & Barnard 79] provide evidence for the benefit of using an establishing, or long shot, prior to using close-ups. A study was performed in which picture-only instructions either showed the context of the object, or just showed a close-up. It was found that performance was poor in a close-up only condition. [Marcel & Barnard 79] conclude 'instructions which show only the relevant parts are more difficult to comprehend and carry out.' They suggest three reasons for this : 'the visual frame of reference in the part instructions may be inadequate, making it difficult to identify individual components'; 'depiction of the entire equipment enables people to better grasp the relationship between parts of the equipment'; 'ambiguities leading false inferences were reduced by visual context'.

[May & Barnard 95] have suggested that similar problems may exist in computer interfaces that zoom, or show multiple views of the same information. They note that a tourist information system which showed two views of a street map caused problems for users because they found it difficult to integrate the different screens when a zoom was made. They suggest that when a new view of the map is shown, all the inbetween frames of a zoom should be shown, and that an element selected to be zoomed should stay in the same place on the screen when it enlarged. Their arguments for continuity follow those of [Hochberg 86]; [Kraft 87].

Summary

- Composition information will initially be extracted from the image at scene level. By default, very little composition detail will be recalled from the scene.

- If an object is important then it should be given emphasis, by making it out of place in the scene. This can be done by adding a symbol to it which has a learnt meaning, or using a highlight which makes it stand out from it's surroundings

- Composition information extracted from an image may be controlled by varying the frame size e.g. using a close-up or establishing shot.

- Caveat : If the composition information is complex or unusual then the whole of the zoom effect should be shown with the image gradually changing size to show the new view.

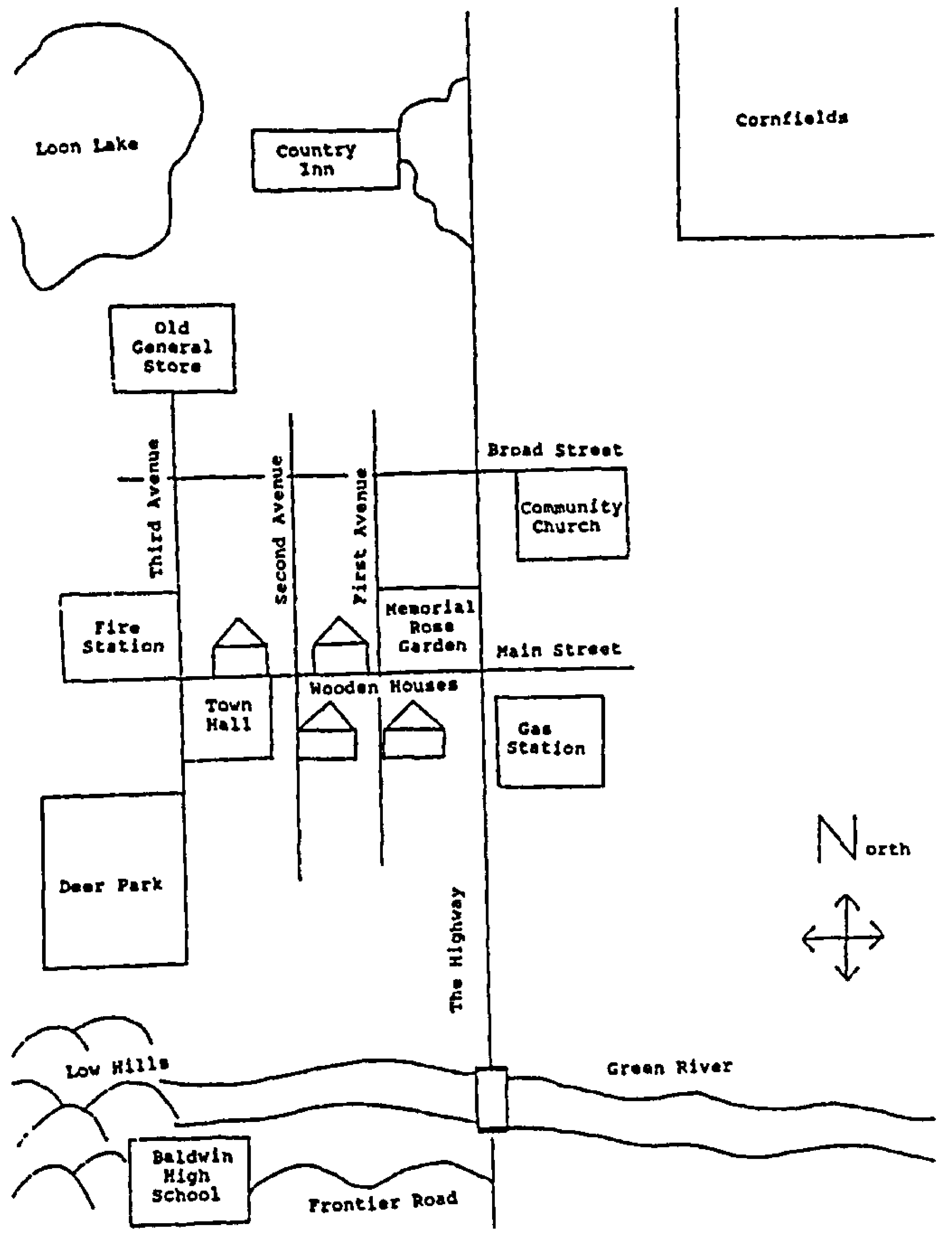
2.7.1.4 Media Selection & Combination : Spatial information

Several studies have shown that pictures are useful for the comprehension of spatial information. [Purnell & Solman 91] conducted several studies using geographic materials describing location of objects given in either text or image. It was found that 'comprehension of spatial content presented as an illustration was found to be superior to comprehension of the same content presented as text.'

[Bieger & Glock 84] also noted improvements in performance of a construction task if spatial information was provided in pictures : 'Groups receiving the spatial information in text performed the assembly with fewer errors, and those seeing the spatial information in pictures completed the task in considerably less time.'

However, [Levie & Lentz 90]; [Dean & Kulhavy 81] both report studies which suggest that a combination of text and image is more useful than image alone in providing spatial information. [Dean & Kulhavy 81] report a study in which subjects read prose describing the location of an African tribe, either with or without an accompanying map. Students who saw the map learnt about 60% more than students who did not. [Dean & Kulhavy 81] noted that if the text did not describe spatial relations, learning fell to a similar level as to that of students who did not see the map. They conclude : 'the map was a significant help only when processing the information on the map was assured.'

Similar findings are cited in [Levie & Lentz 90] in study performed of children's recall of a short study with pictures that either illustrated spatial relations within the story or did not. They found that children who read the spatial illustrated version recalled 'about four times as many spatial relations'.



The little town of Crestview is a mining town. The road to Crestview runs north. Crestview begins where the highway crosses the green river. The river flows to the east out of some hills that lie to the west. Just across, to the north, Crestview high school lies on the bank to the west at the base of the hills.

Figure 2.7.9 Materials used by [Ferguson & Hegarty 94] : text aided recall of the map.

These results suggest that spatial information is best recalled if it is presented as an image. However, the results of [Ferguson & Hegarty 94] also suggest that language may be useful in structuring spatial information.

[Ferguson & Hegarty 94] investigated how providing both a spatial descriptive text alongside an image effect recall. Two main benefits are suggested to having both a visual and text based representation of spatial information (see figure 2.7.9). Principally it is argued that the image displays all the spatial relationships between landmarks and other objects, whereas a text only describes only a portion explicitly and requiring the rest to be inferred. They also noted that the image provides metric information concerning object location which would not be available from the text. These results suggest that text is useful in organising spatial arrangements, and providing particular objects a landmarks. The image itself will then give detail of object arrangements and location.

Whilst [Ferguson & Hegarty 94] studied only the use of text to structure the spatial information in an image, it is likely that a voice over would have the same effect, locating particular objects and setting relations between them.

Summary

- If spatial information is required then prefer an image; an image allows more spatial relations to be recalled than text.

  - Caveat : If a complex spatial arrangement of objects is needed, language can be used to add structure to a set of spatial representations and identify landmarks in a complex image.

2.7.1.5 Within media design : Spatial information in Images

Designing for spatial information concerns how objects will be located, and how certain objects can be selected to be used as landmarks, around which other objects will be organised.

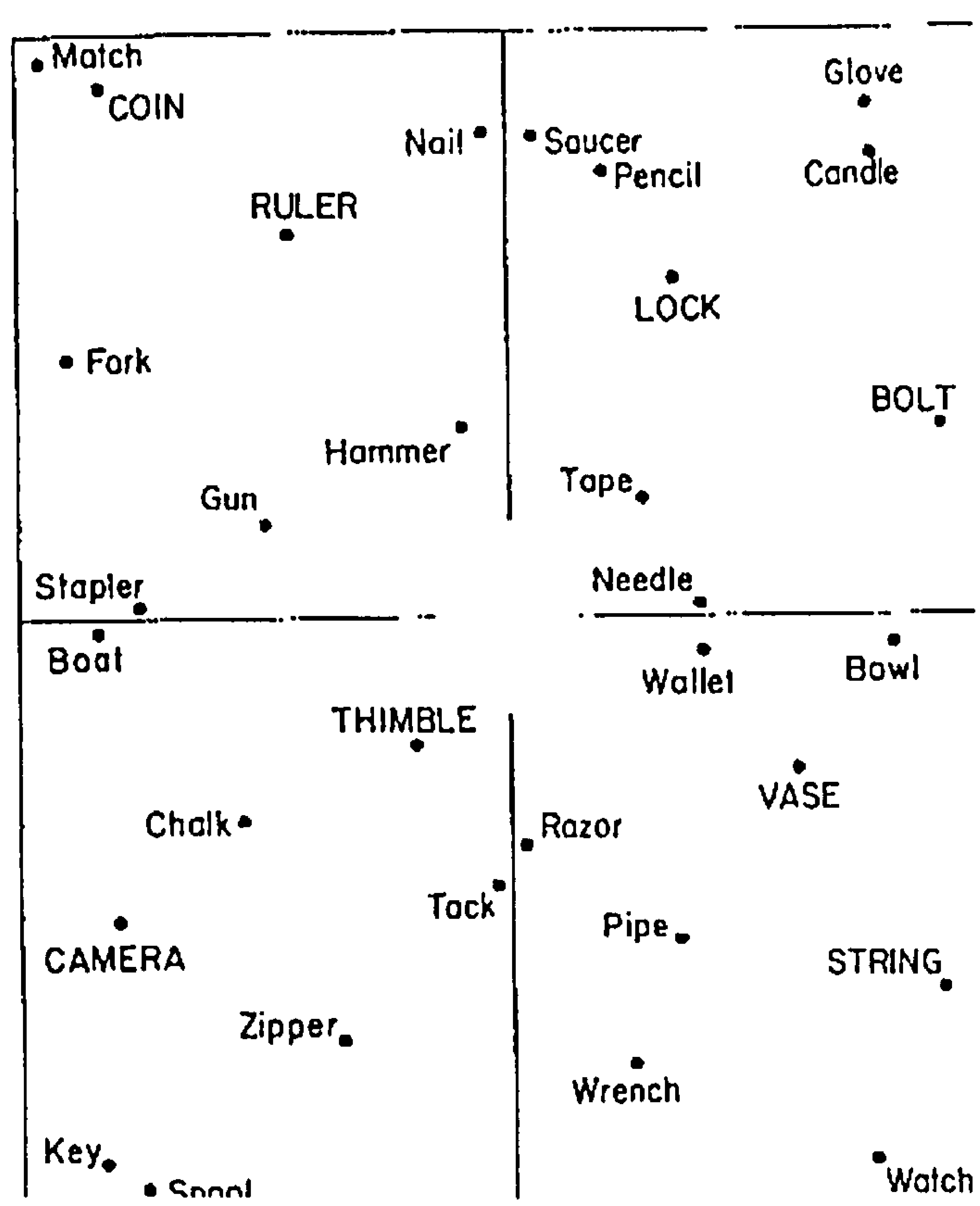


Figure 2.7.10 Materials used by [Macnamara 91]. Objects were used instead of names in the study. When asked to recall the arrangement, subjects recalled certain landmark objects (shown in bold type), then placed other objects in relation to them .

[Mandler & Ritchey 77] provide a general study of recall of spatial information from scene. They suggest that spatial recall could be separated into metric spatial location, accurately locating an object; or relative spatial location, locating an object by a category or relation. It was found recognition of relative spatial information was generally better than for metric.

They also note that in disorganised scenes ie those in which the arrangement of objects did not follow real world organisations, metric information was better recalled

[Macnamara 91] provides a more detailed account of spatial representation taken from images. Two main types of spatial memory are suggested : spatial hierarchies and landmarks. [Macnamara 91] suggests that spatial locations may be grouped together into a hierarchy of regions. A study found that accuracy of object location was effected by priming of nearby locations (see figure 2.7.10). By specifying the regions physically within the picture (dividing it into quarters), it was found that the object locations in the regions encoded were distorted; by underestimation of distance within a region, and overestimation between regions. [Macnamara 91] concludes that these effects may be due to 'two kinds of mental representations' being formed 'a hierarchical structure which is stored in categorical spatial memory, and a metric spatial structure which is stored in co-ordinate spatial memory.

[Macnamara 91] has also proposed that spatial memory may make use of gross spatial relationships between less well fixed or unimportant objects, whilst fixed or important 'landmark' or reference point locations may be encoded using metric information. It was noted that landmark objects were treated differently from other objects : non-landmarks were primed more effectively by landmarks and non-landmarks tended to be placed nearer to landmarks. The study performed by [Macnamara 91] also demonstrated that higher level landmarks could be formed by dividing the image itself into distinct areas. In the case of the study, the image was quartered (see figure 2.7.10).

These results suggest that it is important to select certain key objects as landmarks. [Bryant & Subbiah 94] offer several heuristics for landmark or reference point selection. They suggest that the first means may be due to the physical features and structure of the figure itself. Landmarks may be chosen from 'perceptually salient features'. One source of landmarks may be biased towards objects that are within the focus of attention : 'if attention is focused on a specific part of a figure, features in that region will become more salient as landmarks than features in other parts of the figure'. The area attended to may give a 'frame of reference' from which to encode other relationships and locations.

Summary

- If an image is well organised then the location information extracted will contain only a general category of spatial relations. This may give good recognition of general spatial relationships, but a poor recognition of spatial location.

- If a particular object needs to be located accurately, then set it as a landmark. Landmarks will tend to be perceptually salient objects, or objects with which the user is already familiar and knows the location of. These will be located more accurately and will help organise the location of other objects.

- If several objects are to be located as landmarks, it is important to divide the image into sub-areas if possible and then select a landmark for each area.

2.7.2 Operational information

Operational information concerns action upon objects. It is split into two basic types : Role and Physical Action. Role information concerns the object(s) which take part in the action, such as which object is the agent of the action, and which object is being acted on as the patient. Physical Action information concerns the change or motion which takes place upon the object.

2.7.2.1 Media Selection & Combination : Role

Role information concerns how objects are related within an action. This includes information such as which object is the actor, and which are being acted upon.

Few empirical studies have directly addressed how communicating role effects identification of physical actions. [Michotte 63] suggest that role identification from an animation will not be directly available from the motion; it requires inference. A study by [Michotte 63] has suggested that identifying role may be based on inferences taken from the motion itself. A group of objects are seen to move around an open rectangle - viewers later report that the blocks were 'chasing each other' and 'being hit', with the roles being inferred from the motion (see figure 2.7.11). [Michotte 63] notes that where actions are complex or unusual, these inferences were often found to be wrong.

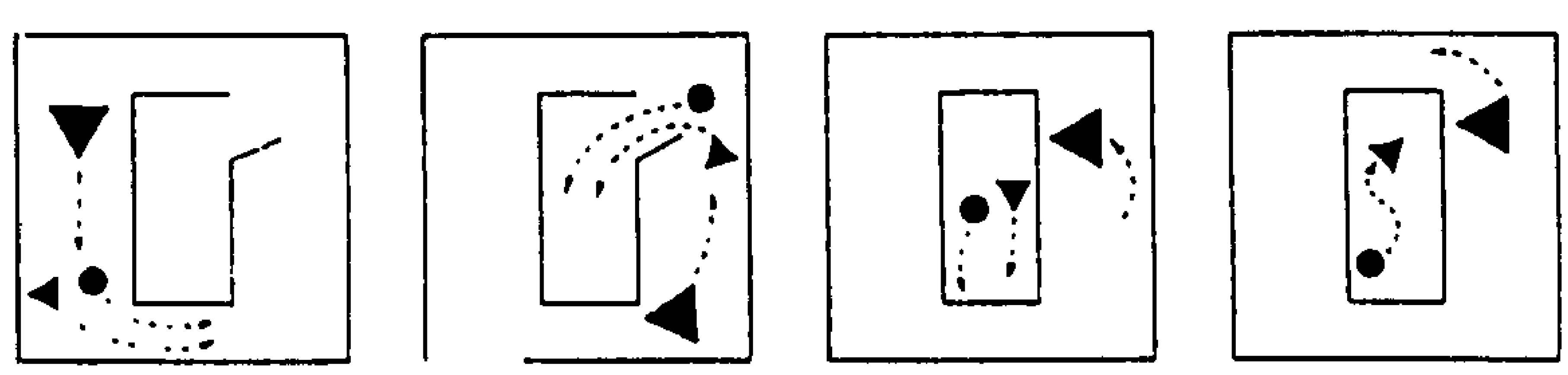


Figure 2.7.11 Materials used by [Michotte 63]. Subjects used inference to identify the role of each object (eg ‘chasing’) , often incorrectly.

[Thibadeau 86] provides an AI based account of how actions may be recognised. He proposed that the representation of an action was based on a set of conceptual procedures, which were triggered due to particular conditions in the object motion. The procedures begin by identifying the object roles within the motion. [Thibadeau 86] suggests that if language identifying the action was available with the animation, it would aid in identifying object roles, since role information which otherwise would require inferences from the animation. The work of [Thibadeau 86], whilst computational in nature, suggests that role information may be best provided in language.

Summary

- In general, to extract role information from an animation or image sequence will require inference; unless the action is well known, it is best to provide role information via language.

2.7.2.2 Media selection & Combination : Physical Action

Physical action information concerns the motion and path which an object follows. It can either be shown as an animation, or by showing a single image, capturing part of the path of the object motion.

[Park & Hopkins 93] offer a review of over twenty five studies of the use of animation, and note that comprehension is improved for a range of tasks that require the understanding of action. They suggest that animation is particularly useful for illustrating a task with many component actions and reactions which can be sequenced in animation; a task which requires movement in several directions at once which would be difficult to process in parallel without animation; or actions which are difficult to describe verbally. [Park & Hopkins 93] conclude that 'concepts involving a time dependent process or trajectory can be represented concretely with graphical animation'. For simpler sequences, [Park & Hopkins 93] suggest that animation is likely to be unnecessary and still images would be sufficient.

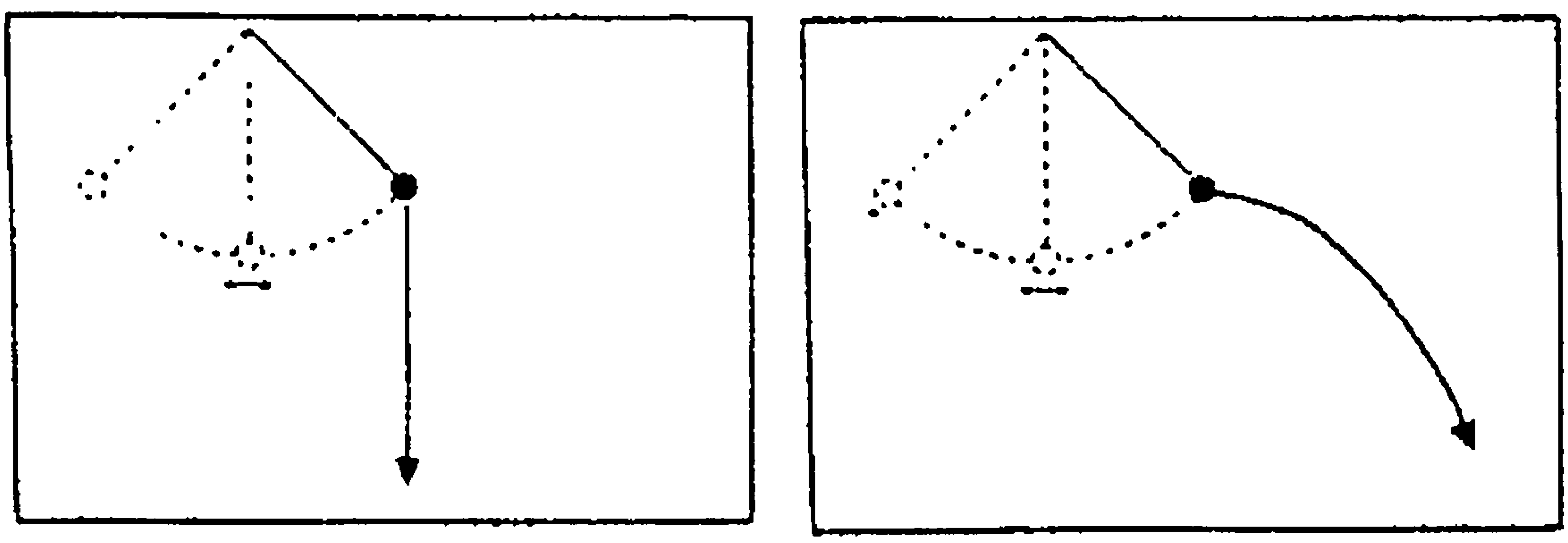


Figure 2.7.12 Materials used by [Kaiser, Proffitt, Whelan & Hech 92] : ‘what happens when the string is cut on a pendulum’; (left) correct answer, at it’s apex the bob has no horizontal velocity so it falls straight to the ground, (right) ‘naive’ answer.

[Kaiser, Proffitt, Whelan & Hech 92] have studied the identification of complex motion from either animation or a single still image. Subjects were given problems involving object motion in which the transition path and breakpoints had to be deduced eg ‘what happens when the string is cut on a pendulum’ (termed 'naive physics' see figure 2.7.12). The use of animation was found to lead to a greater accuracy in the formation of a correct inference as to the type of transition path an object would follow in a particular system. They conclude that animation gives greater motion information than a still image : animation 'segregates in time changes in the dimensionality of an object's motion.'

[Spagenberg 73] investigated the assembly of a machine gun, shown either as a sequence of still images, or as animation. He found that performance was improved if animation was used : 'The results of the two reported experiments indicate a superior performance in the disassembly of a complex weapon by groups seeing a television display which portrays motion than by groups seeing a television display composed of a sequence of still shots.' In particular it is noted that motion is best used if simultaneous motion is required, or if the action is difficult to describe verbally. It was also found that for simple actions, text may be used 'for operations involving simple motion actions which are readily expressed in words such as grasping or lifting, there seems little difference whether motion displays are used or not.' These results agree with those of [Park & Hopkins 93] that animation is only of value for complex motion.

A study of performed by [Swezey 91] provides similar results, based on the performance of a maintenance task with either still images or motion video. They conclude that motion is useful for several conditions : if motion is a defining characteristic of the concept to be taught; the activity requires simultaneous motion in more than one direction; the activity is unfamiliar to the learner; or if the activity is not readily described in words. In other condition still images were found to be as effective as video.

These results suggest that still images may be useful if the action being shown is not complex or unusual. Where the action is more complex, e.g. requiring motion in several directions or complex reactions, animation should be preferred [Swezey 91]; [Park & Hopkins 93].

A study of the use of language in supporting physical action information is provided by [Boohrer 75]. [Boohrer 75] investigated whether language can aid in the identification of motion, using studies of performance in mechanical troubleshooting. He noted that text was useful in reducing errors in the performance of motor tasks : 'the human processing system is most efficient in comprehension of instructions when the pictorial mode is used to aid in selection and organisation of perceptual-motor actions and the verbal material is available to confirm specific actions within the range.' [Boohrer 75] further concludes that 'as more difficult series of actions are required, the verbal channel probably becomes more important' These results suggest that language can also be useful in identifying motion. If the motion is unusual, then animation should be used instead of still images

Summary

- If physical action information is complex e.g. requiring motion in several directions or complex reactions, then animation should be used.

- If the physical action is simple, then their is little difference in using still image or animation media.

  - Caveat : Language may also be useful in identifying a series or group of actions to be understood; or in setting the range of actions to be performed e.g. qualifying the action.

2.7.2.3 Within media design : Physical Action information in Still Images

A still slide may either show an object in transition, or it may show the object at the start or end of its path, termed by [Newtson & Enquist 76] as a 'breakpoint'. The difference in motion understanding produced by this choice between transition and start / end breakpoint in still-motion images has been investigated by [Newtson & Enquist 76]. A set of slides was used presenting break-points or non-breakpoints showing a man attempting to repair a radio. They found that the breakpoints gave a superior sense of order : 'breakpoint slides were judged more likely to portray a caused event as opposed to a random movement than non-breakpoint slides.' Their results suggest that if an action is to be shown as either a single image, or a set of still images, then breakpoints should be shown, i.e. when the motion changes in direction or speed.

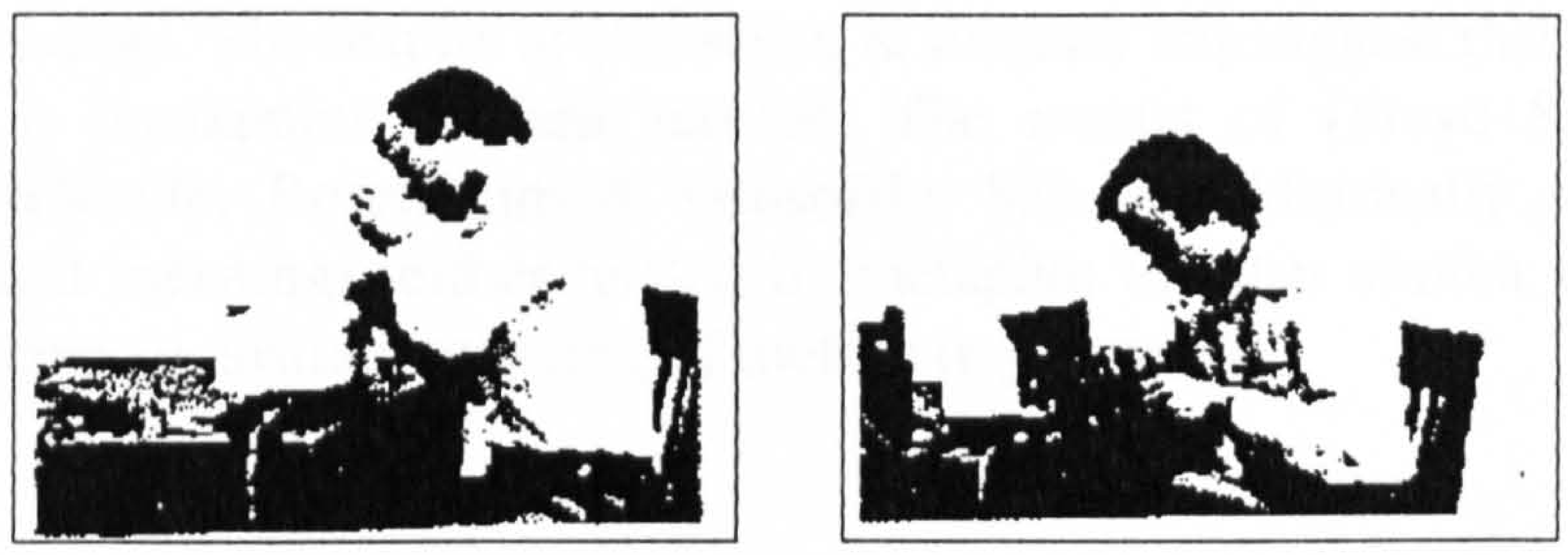


Figure 2.7.13 Materials used by [Newtson & Enquist 76]. The slides show a snap-shot of a transition on the left, with the man's arm in mid-motion; against a breakpoint on the right with the man's arm at the end of the action. The study showed actions were better understood if illustrated by breakpoints.

Several studies have shown that the use of symbols can aid in the identification of motion from a still image [Freyd 83]; [Friedman & Stevenson 80]; [Carello, Rosenblum & Grosofsky 86]. [Freyd 83] suggests that motion is implicit within images that show an object away from its rest position, which is termed 'representational momentum'. She suggests that this momentum is emphasised by the use of symbols such as arrows or speed lines, which lead to a particular direction of change being extracted from the image. A study is reported in which subjects recalled objects as being further along their paths if they were shown with arrows.

[Friedman & Stevenson 80] provide an analysis of the types of image design and symbol which lead to motion being inferred. They suggest the following forms of representation may be used to depict motion in still images :

- Single viewpoint, single moment : show object away from its rest position
- Metaphor : 'Aspects of the environment which are unlikely to occur together in the real world may be represented side by side in a picture so as to suggest movement (e.g. motion blur)';
- Abstract representation : a symbol has a learnt meaning of motion or change (e.g. arrows)

Unfortunately, the suggestions of [Friedman & Stevenson 80] were not based on any empirical evaluation and the different techniques are not rated for effectiveness.

[Carello, Rosenblum & Grosofsky 86] performed an empirical study of the value of pictorial devices. They compared a number of pictorial devices designed to show motion. They note that several pictorial representations can show motion : action lines, multiple viewpoints, postural deviation from rest, inclined orientation, and location of the ground. They suggest that symbols such as arrows and speed lines aid identification of motion because they show the change in motion over time : 'if the event to be depicted is movement in general then a device that captures the path of movement ought to be effective.' They asked subjects to rate the effectiveness of the symbols in showing the action of a man running, jumping or moving.

Results showed that effectiveness of a particular symbol depended upon 'whether a given device highlighted a distinctive aspect of the particular event'; 'if action lines illustrate less important aspects of an event then no one will rely upon them.'

These studies suggest that image design and a number of pictorial symbols can be used to show motion or change. The results of [Newtson & Enquist 76] suggest that objects should be shown away at a breakpoint in their motion. The results of [Freyd 83]; [Friedman & Stevenson 80]; [Carello, Rosenblum & Grosofsky 86] all additionally note the value of symbols with learnt meanings, either related by metaphor to other similar events (e.g. speed lines), or which have an abstract meaning of motion (e.g. arrows).

Summary

- For physical action information given in a sequence of still images it is important that breakpoints are shown as the start and end points of the action, and any changes in direction or speed are shown within the image sequence for the action to be identified.

- Pictorial symbols may aid in the identification of physical action information. These should be chosen to make the style or path of motion more important.

2.7.2.4 Within media design : Physical Action information in Animation

Design decisions for the use of animation include which part of the motion to show, how to use cuts to remove parts of the motion, the use of slow motion to aid in the identification of the motion.

[Brewer & Dupree 83] provide a study which suggests that details of action from animation alone is poorly recalled. A study investigated the recall of actions from a video animation showing a simple set of actions being performed, such as standing on a desk to remove a light bulb. They found within a short period of time memory for the internal details of the action will also be lost, over longer periods any actions not part of a more general goal, will also be lost : 'immediately after viewing an action the observer converts the actions into larger goals'. This process is given to be cyclic, thus as the event unfolds lower level action information will be further integrated and subsequently lost : 'most of the specific physical information (such as right 50cm) is lost the representation immediately.'

The conclusions of [Brewer & Dupree 83] suggest that actions were grouped and surface detail lost were found with actions which were shown as a whole, without any design effects. [Koopman & Newtson 87]; [Newtson, Rinder, Miller & Lacross 78]; [Kaiser, Proffitt, Whelan & Hech 92] investigated how techniques such as slow motion and cutting can effect recall of actions.

[Koopman & Newtson 87] suggest that the 'size' of action encoded will dictate the type of information available subsequently at recall or subsequent performance. [Newtson & Ridner 79] investigated how ‘breakpoints’, or changes in the direction of motion, were extracted when a video was shown in slow motion. They found that a slow motion play back of an event was found to produce better recall of lower level actions.

[Hochberg 86] suggests that using a 'cut' in which part of the motion is removed, can also effect how an action is recalled. [Hochberg 86] suggests 'cuts are used to provide joints at which one performance or enactment of an event can be replaced by another', and that their effect is that 'the viewer is not required to sit through periods during which attention is neither required nor rewarded.' Figure 2.7.14 shows how a sequence can have parts of actions removed eg of the man getting out of the chair and opening the door. [Hochberg 86] also noted that a cut will increase the detail in which actions are encoded following the cut, a cut can thus also alert the viewer to prepare to attend to a new action.

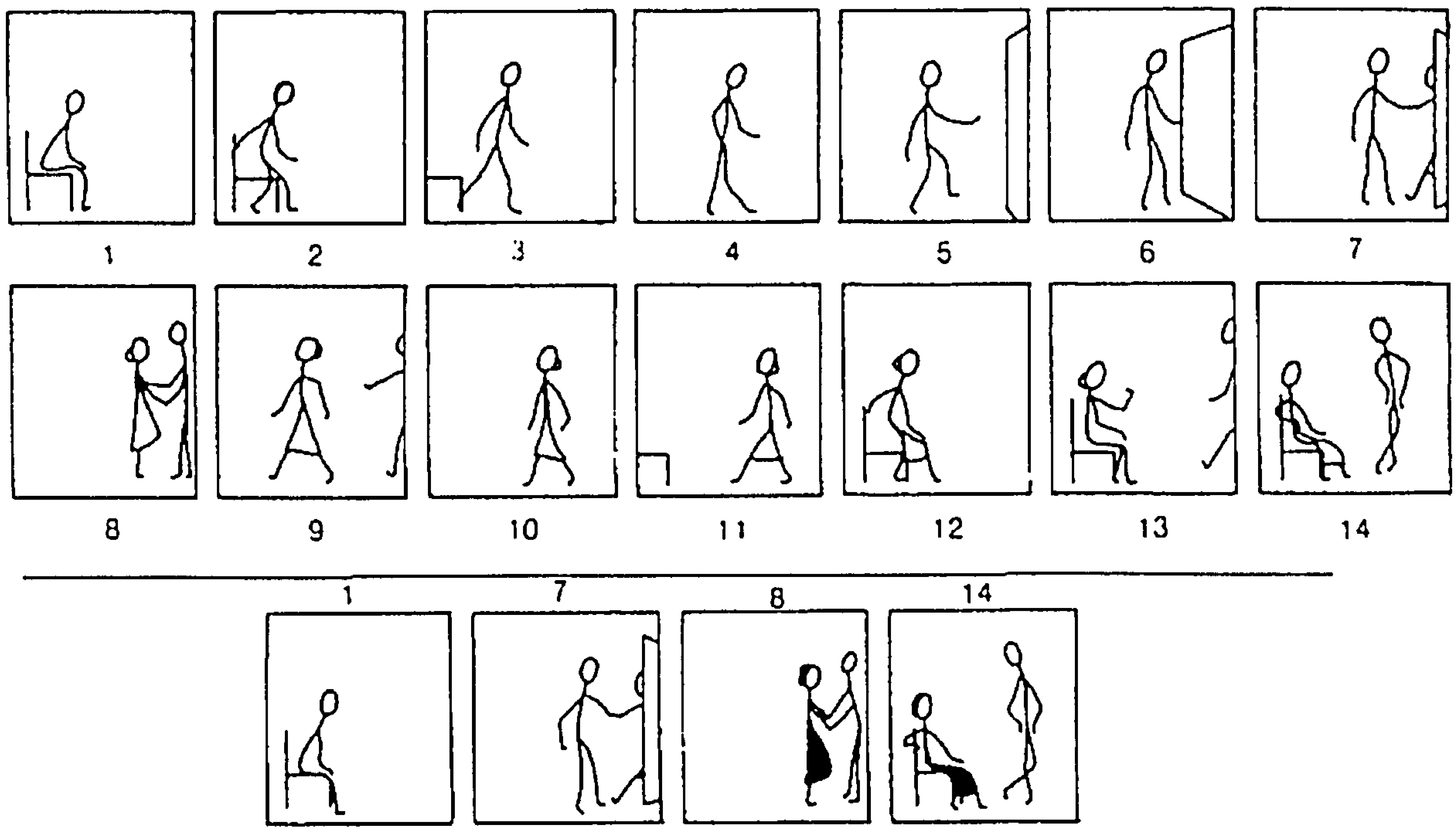


Figure 2.7.14 Example of effects of cutting from [Hochberg 86] : the original sequence of frames 1-14 are cut to frames 1,7,8 & 14 to remove parts of the action eg walking to the door.

[Newtson, Rinder, Miller & Lacross 78] found that where the actions contain cuts, action analysis will be toward a fine level of granularity. An experiment was performed with a video showing paper being stacked. They found if a cut was applied so that the stack was not shown to the viewer then action analysis remained at a finer level. In addition, if an unexpected action was cut in part way through the sequence then the level of action analysis was also increased. They suggest that the analysis of actions may thus start of at a fine level detail and build up to a higher level if the action unfolds in an ordered fashion; however if anything unusual occurs then the analysis will be reduced again and sampling will increase as the viewer attempts to understand the action. They also note that if the cut removes to much of the action e.g. the whole of a reaction, then it will confuse the viewer. These results are similar to those of [Newtson & Enquist 76] who suggest that certain 'breakpoints' in the action are used for identification; removing them would make encoding the action more difficult.

These studies suggest that cuts and slow motion provide a useful way of structuring the detail which will be recalled from an animation. Using slow motion will improve recall of detail; using cuts can remove unnecessary actions and also improve recall of actions shown.

Summary

- Recall of details of action from animation alone is generally poor; animation design techniques such as slow motion and cuts will improve memory of the action.

- Caveat : If the task requires that the physical action information is grouped as a more complex action, the breakpoints of the action should be shown without cuts.

2.7.2.5 Within media design : Role information in Animation

Role information in animation requires that each view of the object must be matched to the previous view. The design of animations to aid this matching process has been investigated by [Hochberg 86]; [Kraft 87].

[Hochberg 86] has suggested that the identity of an object in motion is maintained by a set of correspondence rules. Three distinct phases are given for identifying the object in motion. The first phase is termed 'fast, stimulus driven' matching, which is insensitive to detail, producing a match if gross object detail is the same between frames; it occurs rapidly and automatically. If this fails, then the second phase is 'slow, stimulus driven' processing, based on matching shape found by the first process. Finally, 'slow, cognitive driven' matching is used, by which the viewer may use inference to match the object between frames. [Hochberg 86] performed several studies which made use of abstract patterns which had to be matched, noting that the difference in accuracy of match between the images corresponded to response time in accepting or rejecting them as part of the same overall pattern.

The rules suggested by [Hochberg 86] are expanded by [Kraft 87] who provides a further set of guidelines based on film craft. Unlike [Hochberg 86], these rules are proposed as concerning real world image sequences. In order for motion to be grouped into a single action several continuity rules must be obeyed. These ensure that the two sequences can be matched together into a single action :

- Temporal continuity : action should be presented smoothly; gaps must be explained by cover shots
- Subject continuity : when cutting on the same subject, there should be sufficient communality between shots to permit identification of the subject
- Vector continuity : stationary features that guide eye movements should match up between shots

[Kraft 87] conducted a study using a film sequence in which continuity was either maintained or violated. Kraft notes that 'violating directional continuity disrupted viewers expectations concerning cinematic space', and suggests that such violations reduced comprehension of the action being shown. He concludes that 'viewers could not properly orient the characters in the scene'.

The results of [Hochberg 86]; extended by [Kraft 87] to real world images suggest that it is important that the appearance of the agent and ground are kept constant. [Kraft 87] notes the value of this as 'continuity', allowing several views of a single object to be matched together into a single path of motion.

Summary

- For role information in animation, the appearance of the agent should be identifiable throughout the action and the ground over which the action takes place should be held steady and identifiable.

- If the agent is changing location, it is important that it can be recognised as the same identity. This can be supported by ensuring that the object has a constant appearance during its motion.

2.7.3 Procedural information

Procedural information concerns sequencing actions, such as an action 'follows', is 'concurrent with', 'next'.

2.7.3.1 Media selection & Combination : Procedural information

The effect of providing procedural information in text against the animation has received little direct empirical study. [Palmiter et al 91] notes the value of text in providing sequencing information has also been noted in studies, but has not been directly tested

In animations, temporal order is given by using cuts, or wipes and dissolves in which the screen is faded to black. [Burch 73] has noted that in film theory, the use of these techniques does not directly provide any notion of time frame or sequenced order. It is suggested that captions may be useful in locating a particular point in time (eg' In the year..'), or in sequencing (e.g. previously...'). However, no empirical support of the value of these text cues is provided.

A more useful study of the value of providing procedural information in text is provided by [Palmiter et al 91]. She compared the recall of procedural information either given in animation, or animation with text in the performance of a computer based tutorial. She suggests that text is useful in breaking instruction into steps, with animation supporting each step within the task. [Palmiter et al 91] notes that 'Instructions need to logically segment a task, especially for novice users.'; 'each step of a textual procedure, containing an action and object would provide segments.'

Summary

- Language should be used to structure procedural information, particularly when the time frame is not constant; animation or image sequences will be useful to support the underlying physical action information.

2.7.3.2 Within media design : Procedural information in Text

In text, procedures are signalled using cue phrases. This has been studied by [Gernsbacher 85], and in more detail by [Betsgen & Costermans 94].

[Gernsbacher 85] found that by inserting 'next' between a series of related propositions, information from the preceding sentences was found to be less available to recall. A sentence was presented such as 'The lifeguard was watching the children swim. He noticed one child was struggling. He thought the child might down', followed by either 'Next, he jumped into the water. He began to administer CPR' or 'He jumped into the water. Next he began to administer CPR'. [Gernsbacher 85] found that the sentence which began with 'next' was read more slowly, and that recall of information from previous sentences fell.

[Betsgen & Costermans 94] have expanded the initial study of [Gernsbacher 85] and provided more detailed results. They suggest that temporal relations in language can be divided into two classes : anchorage markers and sequence markers. They analysed how subjects organise a list of related statements which contained temporal cues. Anchorage markers indicate a particular point it time; whilst sequence markers place the event in relation to another event within the same period time.

Another way of communicating procedural information in text is to use formatting or instructional graphics. [Armbuster, Anderson & Mayer 91] suggest a variety of techniques which can be used to provide organisation for text sequences, such as lists and frames separating each procedure. A study by [Armbuster, Anderson & Mayer 91] found that flow charts and lists were useful in organising sequences within text : 'frames help readers with at least two of the basic cognitive processes in reading : selecting and organising information from text', 'frames show the logical and temporal relation between ideas in a text, they can help build internal connections between important text ideas.'

Summary

- Procedure sequencing information can be given in language by the use of anchors or cue phrases

- A cue phrase (e.g. 'at time x') should be used to locate a particular important time point in the task, such as the start or end of a sequence.

- Cue phrases can then be used to order the sequence ('next', 'then').

- Structure text to indicate procedure sequencing by formatting into lists or frames.

2.7.3.3 Within media design : Procedural information in Animation

Animations may be segmented either by a change from one event to another by a cut, or by a dissolve. Cuts will generally be less well recognised within the animation [Kraft 86] and are generally used to join order events into a linear sequence. These would be used for smaller scale time changes. Dissolves are used to provide a more distinct change in the animation, fading the screen to black. These are used to indicate a greater change of time, which would be over a greater period of time. The most well developed theory of how temporal models are built between scenes or actions can be taken from film theory. [Burch 73] suggests that an animation can contain five types of temporal articulation between two sequences :

- Absolutely continuous
- Time ellipsis (time abridgement - part of action omitted) - requires some spatial continuity to 'measure' the extent of the omission.
- Indefinite ellipsis - not measurable (flash forward)
- Time reversal (repeating part of action or reaction for effect e.g. opening door, then showing room before door opened) - overlapping cut (may also be required for a matching cut to make it appear smooth)
- Indefinite time reversal (flashback)

[Burch 73] offers the types of articulation based on rules of thumb and from analysis of existing films. He does not provide any empirical support for their effectiveness.

Empirical evidence for the comprehension of action into a sequence when separated by cuts is given by [Carroll & Bever 93]; [Gernsbacher 85]. [Carroll & Bever 93] conducted a study of recall of actions from a film which were divided by either a change in action or a cut, or by both. It was found that recall of the action after the cut or change in action was superior to that prior to it 'indicating that the material in the first segment is organised separately and more abstractly than the material in the second segment'. They conclude that 'viewers initially process material in real time until they reach a structural boundary. At that point all material is recoded into a more abstract and holistic form.'

Figure 2.7.15 Example of materials used by [Gernsbacher 85]. Recall of details from the first four frames was worse when the fifth frame was shown, which is a structural boundary of the sequence.

[Gernsbacher 85] has also conducted studies which investigated both the recall of sequences of action shown as a cartoon strip (see figure 2.7.15). The results suggest that actions are segmented at certain points into a procedure. She suggests that these points are defined when a series of actions can no longer be combined into a single whole, and a new 'structure' of actions must be constructed.[Gernsbacher 85] notes that details of the previous actions were less available to recall following a shift to a new procedure.

These results suggest that using a cut in a sequence of actions will help identify a sequence of actions; with the action following the cut being treated as the start of a new sequence. The more general conclusions of [Madsen 78] suggest that wipe / dissolve filmic techniques may be used if the sequence is separated by a greater period of time, or a cut used if the sequence is nearer in time. A new sequence will also be started if the current actions can not be matched into a single complex action.

Summary

- Procedure sequencing information may be given from two types of filmic technique signals: either a cut, or wipe / dissolve. These may help in emphasising that a change in sequence has taken place in the animation.

- The sequence will be assumed to be continuous unless a cut, wipe or dissolve filmic technique is encountered; and if the actions can be combined into a coherent goal.

- Following a cut, wipe or dissolve descriptive and spatial information will be less available; use these techniques only when necessary to define the sequence.

- If the change in sequence required is large, such as a flash forward, then a wipe or dissolve should be used as an explicit signal. Cuts in general signal a smaller change in time.

2.7.4 Causal

Causal is the high level, cause-effect relations explaining how and why some part of the world behaves.

2.7.4.1 Media Selection & Combination : Causal

Causal information provides higher level structure, relating actions into cause-effect relations in the world. Most empirical findings [Mayer & Anderson 91]; [Carpenter & Just 92]; [Park & Hopkins 93]; [Swezey 91] support the need for text within animations or image sequences which are to provide causal information.

[Mayer & Anderson 91] gave subjects an animated sequence depicting the workings of a brake system (see figure 2.7.16). It was noted that comprehension of animations depicting actions or events which are unfamiliar may be poor due to a failure to link them together with their causality.

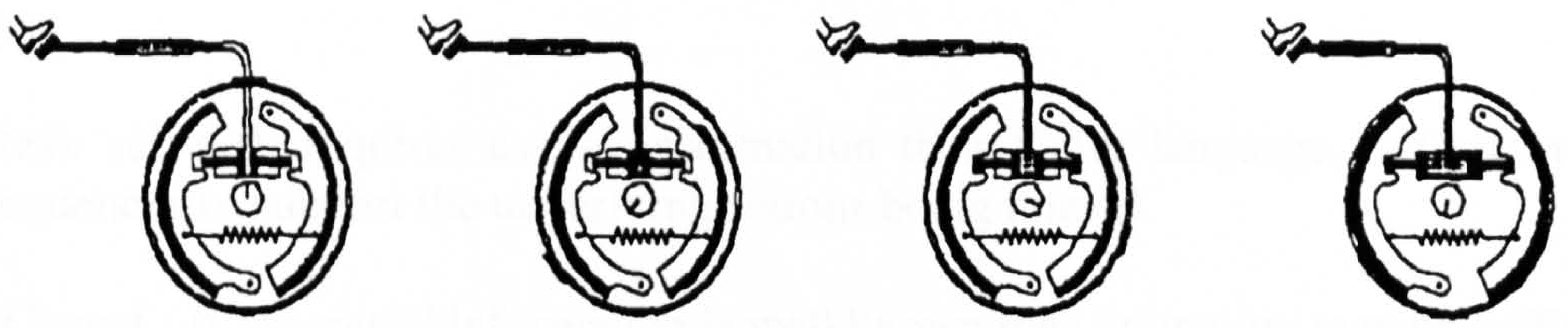


Figure 2.7.16 Materials used by [Mayer & Anderson 91]. Recall of causal details as to how the break operated were poor.

More detailed findings are provided by [Carpenter & Just 92]; [Park & Hopkins 93]; [Swezey 91]. [Carpenter & Just 92] found that text was vital in comprehension of mechanics problems. They argue that subjects who had a low domain knowledge were primarily text dependent, and animation did not provide a good model for identifying the action as part of the system function and structure.

[Swezey 91] performed a study which compared the use of text and still images against text and animation to provide information for a fault finding task which required both a knowledge of causal information, such as system structure and function; as well as the performance of actions. It was found that provision of causal information in text improved subsequent performance on similar tasks over the use image or animation; and that the choice of image or animation did not improve the performance of diagnostic tasks : 'In this study use of visually presented motion during training was not found to enhance either maintenance performance or transfer on a trouble shooting task'.; 'When a transfer task was used as the criterion measure, results indicated that training that included (text based) conceptual information concerning a system's structure and / or function improved performance.'

[Park & Hopkins 93] review [Carpenter & Just 92] and [Swezey 91] and suggest that the design of accompanying verbal information is an important consideration for the use of animation. 'The visual displays, particularly animation, should be presented with sufficient verbal explanation. If the dynamic features are an essential component of the to-be-learned task, either an inadequate explanation or no explanation could result in the desired connections not being processed.' [Park & O'Choon] conclude that if the causal information is simple or well known, then animation may be sufficient, but if it is complex or unfamiliar, then language media are required.

A developmental study performed by [Trabasso & Nickels 92] provides further evidence for the difficulties of extracting causal information from a sequence of images. [Trabasso & Nickels 92] tested children's recall of a simple story shown as a sequence of images, without any accompanying text. It was found that unless children had some background knowledge of the causal relations shown within the images, they would not be able to connect the separate image together. They conclude : 'most 3 and 4 year old children describe the pictures in isolation. The important developmental finding is that from the age of 5 to 12, the children increasingly organised the events in relation to an overall plot line'. These results suggest that if animation alone is shown, the causal information must be already familiar or non-complex..

Summary

- If the task sub-goal requires causal information then prefer language, with animation or images sequences to support the underlying actions being related.

  - Caveat : If the causal information is well known then animation may be sufficient.

2.7.4.2 Within media design : Causal information in Text

The design decisions for text providing causal information is how to structure the text to support the relations required. [Kintsch & Yarbrough 82]; [Meyer, Brandt & Bluth 80] have suggested readers rely upon rhetorical cues in text such as 'because, 'in order to', 'allows' help organise the texts meaning. These are defined more formally by [Meyer, Brandt, Bluth 80] as 'information in text which does not add new content on topic, but which points out aspects of the structure of the content'. These definitions suggest that such cue phrases are organising the dynamic model of the text.

[Meyer, Brandt, Bluth 80] conducted experiments using texts which made use of cue phrases whilst others did not (see figure 2.7.17). They found that if the text did not contain rhetorical cues then the reader was required to use general world knowledge to infer the required causal relationships. Comprehension was found to be poor where knowledge was low. Improvements in comprehension were given if text was well designed to make rhetorical relations in the content explicit They conclude that removing these explicit surface cues reduced comprehension of the passage and led to more errors in subsequent recall.

> <u>THE SOLUTION TO THE PROBLEM</u> IS NOT TO IMMEDIATELY HALT THE USE OF TANKERS ON THE OCEAN since about 80% of the worlds oil is carried by tankers. <u>INSTEAD THE SOLUTION LIES IN THE TRAINING OF OFFICERS OF SUPERTANKERS, BETTER BUILDING OF TANKERS, AND INSTALLING GROUND CONTROL STATIONS TO GUIDE TANKERS NEAR SHORE.</u>

Figure 2.7.17 Example of materials used by [Meyer, Brandt, Bluth 80]. Underlined parts contain cue phrases. Cue phrases aided comprehension.

Titles and headings provide cues to text specifying causal relationships by making them salient from other surrounding text. These are termed 'advanced organisers' and may also help to structure the causal information from a text. [Ausubel 62] defined an 'advanced organiser' as 'appropriately relevant and inclusive introductory materials..introduced in advance of learning.. and presented at a higher level of generality and inclusiveness'. Advanced organisers are shown to before the topic itself is presented, and use typographic techniques such as titles to organise complex causal information.

[Dinnel & Glover 88] have studied the effect of titles, comparing how well a text was understood with or without organisers. They suggest that a title 'helps to provide ideational scaffolding for the stable incorporation and retention of more detailed information that will follow'.

Summary

- If the task requires a set of relations, then an advanced organiser should be used to organise the information.

- Rhetorical cue phrases e.g. 'because, in order to', should be used if the causal information is complex.

2.8 Conclusions

This chapter has presented a framework for perception, attention and comprehension of multimedia, based on a review of literature. The framework provides a number of sets of conclusions :

i) Perception, STM and attentional processing of different media, and combinations of media. Key issues included 'attentional design', how elements of them presentation to be designed to make them more salient, or 'emphasise' them; and what restrictions STM places on how auditory and visual media elements can be presented.

ii) Working memory is used to combine verbal and visual media together. Key issues included how attention and working memory can be best used to bring together speech, text, image and animation to form a single whole; and the use of 'Contact points' in which visual and verbal media are integrated together by co-reference.

iii) Comprehension of multimedia. Issues concerned 'media selection and combination', which media or set of media are best to support comprehension of a particular type of content, and 'within media design', how should the media be designed to aid comprehension of a particular type of content.

The conclusions are used as the basis for a design method described in the next chapter.

Chapter 3

Design Method

3.1 Introduction

In order to be of use to a designer, the framework presented in chapter 2 needs to be re-worked as a method, with a number of steps guiding the designer to ask questions about the multimedia presentation being designed. A synthesis of the conclusions from each section of the framework was used to provide guidelines and to structure the method around the key multimedia design problems. The method was originally reported in [Faraday & Sutcliffe 93] and developed further in [Sutcliffe & Faraday 94].

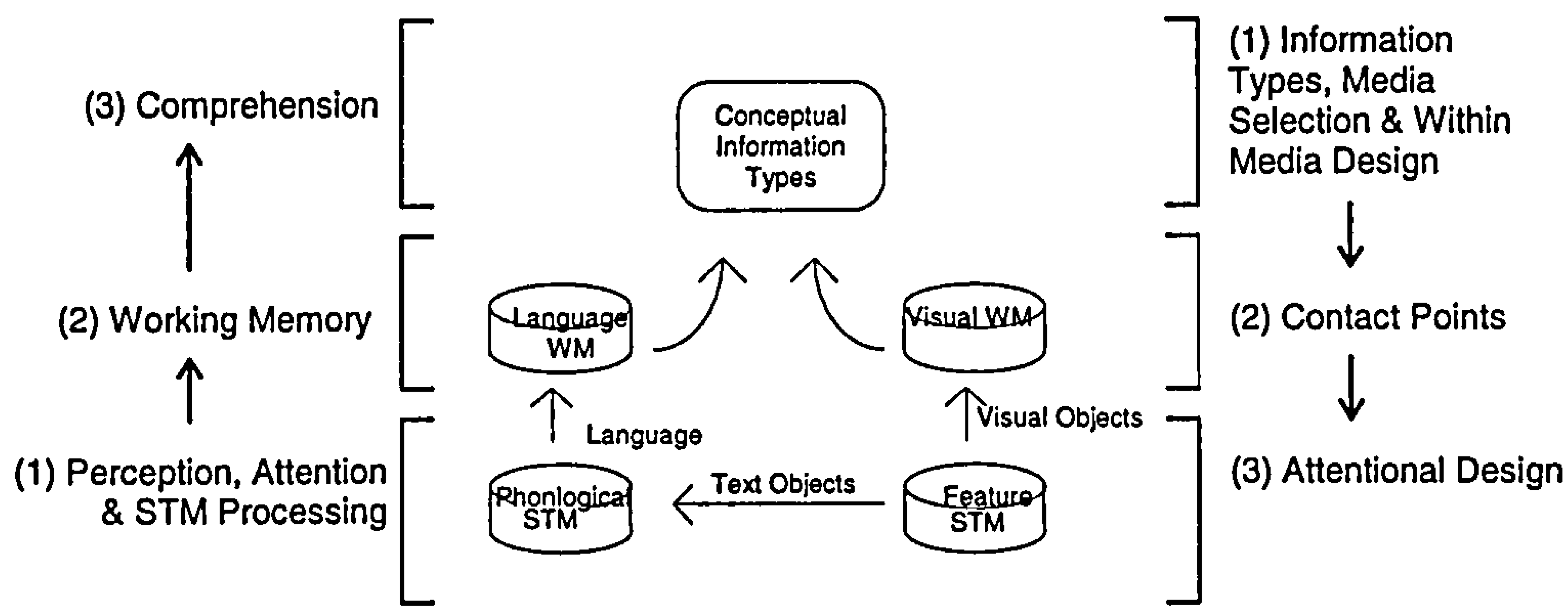


Figure 3.1 Cognitive framework (left-side) and related multimedia design issues (right-side)

The design method inverts the framework given in chapter 2, as shown in figure 3.1. Thus rather than considering the flow of information processing from perception and attention, via working memory, to comprehension. The method starts with choosing the information type to be comprehended, provides advice on media selection and combination, then gives 'within media' design guidelines for each information type, design for working memory and for attention. This gives the following method steps :

i) Design for media selection & combination : which media or combination is best to support a particular information type ?

ii) Within media design : how should the media be designed to aid comprehension of a particular information type ?

iii) Design for contact points : how to bring together speech, text, image and animation to form a single whole; and the use of 'contact points' in which visual and verbal media are integrated together by co-reference ?

iv) Attentional design : how elements of the presentation are designed to make them more salient, or 'emphasise' them; and what restrictions do cognitive resources place on how auditory and visual media elements can be presented ?

The method guidelines are offered as a summary of the studies found in chapter 2, presented in a way which can be used to design a multimedia presentation. It is important that the method should be well structured and motivated by examples of its use. A study of the use of an initial draft of an ISO standard for menu interfaces by [de Souza & Bevan 90] found that 11% of the guidelines were invalidated by designers, and that designers had difficulty in interpreting 30% of the guidelines. [de Souza & Bevan 90] concluded that the way in which guidelines are presented and derived from the underlying theoretic or empirical evidence is of key importance. They suggest that it is vital to provide examples of the guidelines application in the form of case studies showing both good and bad design, and to clearly scope the applicability of the guidelines.

To be of use to a designer, the method thus need to include notions for the models proposed, should be structured in such a way so that the guidelines for each stage easily be referenced, need to be motivated by a case study illustrating its application in good and bad design, and should be clearly scoped.

3.2 Method Overview

The agenda of issues which a method must address were first, the creation of a task model incorporating specification of information requirements and presentational effects. The method should advise on selecting appropriate media for the information needs and scripting a coherent presentation for a task context. Secondly, the design must deal with directing the user's attention to extract required information from a given presentation and focus on the correct level of detail. In addition, the design method should guide the designers to the cognitive issues underlying a multimedia presentation such as selective attention, concurrency and limited cognitive resources such as working memory. Figure 3.2 gives an overview of the method components.

In producing the method, several restrictions were placed upon its scope. These are as follows :

- The method does not offer guidelines for design of interaction, such as how to link between sequences or script user interaction. The key issue for the method was the design of presentations.

- The method does not address the design of the content itself below the level of information types, and addresses only expository materials, such as how-to-do-it style presentations. More complex or abstract information are beyond the remit of the method.

- The method does not address the design of entertainment titles, or presentations which are designed to have an emotional or aesthetic impact. In many cases, these materials produce their effect by confusing the user, or using media in an unconventional manner and are outside the bounds of the method.

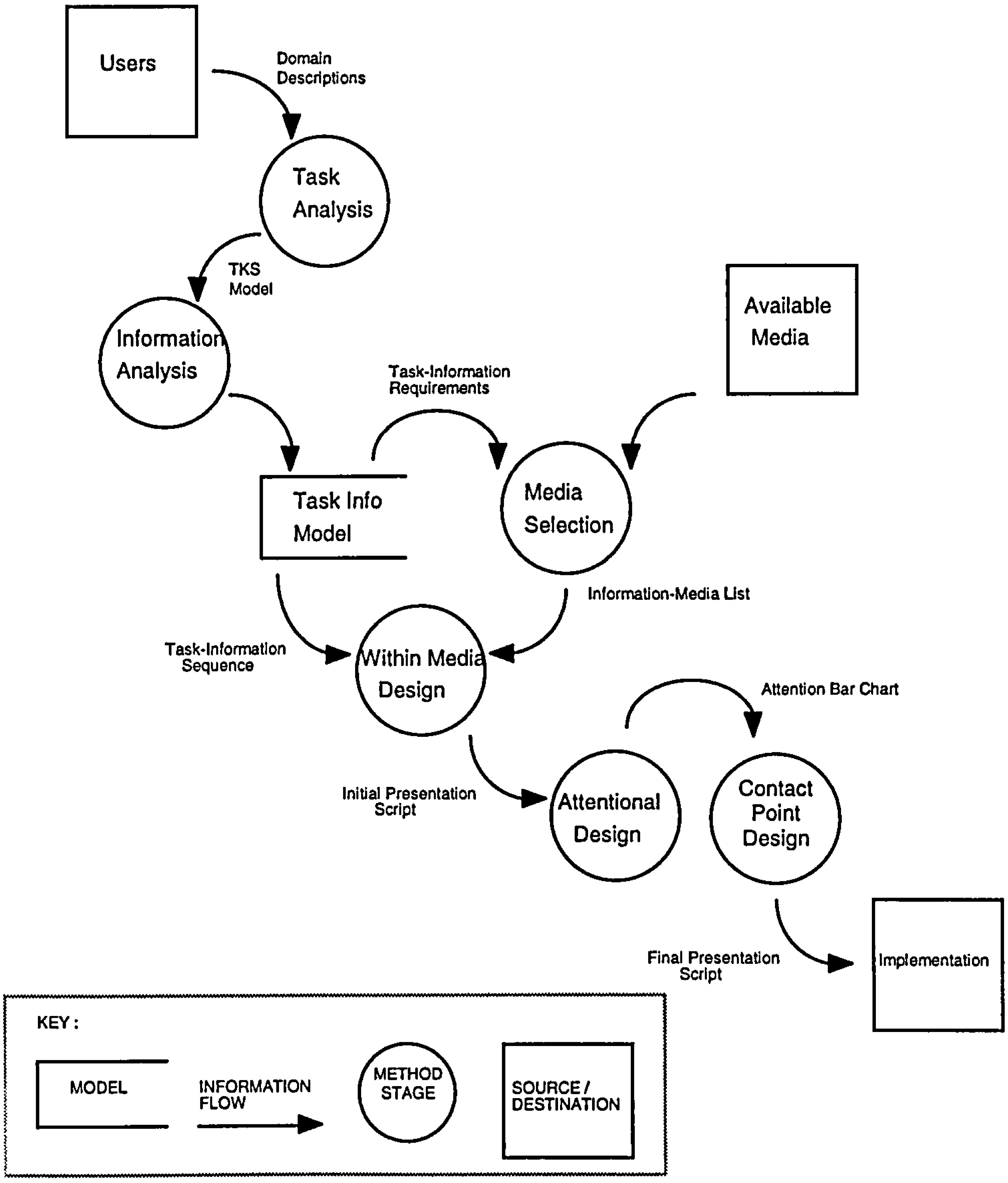


Figure 3.2 Method Overview

The method uses four of sets of rules, each of which are identified by an acronym, shown in table 5.1 :

Method stage	Rules Types	Acronym
Media Selection	Selection Rules	**SR**
	Validation Rules	**VR**
Within Media Design	Image Presentation	**IP**
	Animation Presentation	**AP**
	Linguistic Presentation	**LP**
Attentional Design	Image Attention	**IA**
	Animation Attention	**AA**
	Text Attention	**TA**
	Speech & Sound Attention	**SA**
Contact Point Design	Contact Point	**CP**

Table 5.1 Acronyms used within the Method

3.3 Case Study

The method is illustrated by a case study scenario. It is based upon a set of requirements taken from Lloyds Registry. This concerns the design of a shipboard emergency decision support system to enable a ship's captain to manage hazards such as fire, collision, and chemical spillage. The system's prime role was to provide appropriate information for human decision making, rather than automating the task per se.

The case study is used to illustrate a number of good and bad design alternatives which the guidelines suggest. A final version of the case study presentation is then given in section 3.9 .

3.4 Task Information Model

The method starts with a standard task analysis using Task Knowledge Structures (TKS) [Johnson et al 1988] as the basis for the Task Information Model. A small fragment of the overall task model, the goal tree for managing shipboard emergencies, is illustrated in figure 3.3. The task model is made up of a set of sub-goals, the top level representing an overall goal (e.g. 'Manage Emergency'), the second level representing Procedures required for the goal (e.g. 'Fight Fire') and the leaves representing sub-goals (e.g. 'Muster Team').

To form the Task Information Model, the initial TKS representation must be enriched by adding the information type that each task sub-goal requires, and the presentational effect to focus the user's attention on important information for the task. TKS does not explicitly state what information is required for each task step, although this may be implicit in the object descriptions.

The task model is elaborated by attaching information types, which specify the content to be communicated the user, and presentation effects, to specify the desired attentional effect of presenting the information types. When annotated, the task model should allow the designer to answer the question 'what information content does the user need for this task sub-goal and what important information should the user attend to ?'. The information types are taken from chapter 2, section 2.7, based on [Miller & Johnson-Laird 76].

The following information types are proposed to categorise objects and procedures identified in the task model. The main distinctions drawn are between descriptive/ operational (how to do it) and concrete/abstract information.

Descriptive :

Physical describes an object's physical appearance, e.g. colour, texture, shape, size, etc.

Composition concerns the arrangement, aggregation or composition of an object e.g. whole-part relationships, information that an object is 'part of' another, or that an object 'contains' another.

Spatial information concerns an object's location, including position, placement and orientation; adjacency to other objects e.g. relative position, above beneath; and path.

Spatial differs from composition in that composition concerns aggregation of parts into a whole.

Qualifying information is non-physical, and not immediately visible facts, such as number, object class, status.

Operational :

Physical Action information concerns action which results in change in location or state by the agent, e.g. move left, ascend, follow pathway. This should illustrate how to perform a particular action which 'changes the position' or 'changes the state' of an object.

Role identifies how objects are related with an action. This connects the object effected by an action so a particular object 'is acted upon' as the patient ', or 'acts upon another' as the agent.

Organising :

Procedure is a temporal sequence of actions or states with control operators, such as an action 'follows', is 'concurrent with', 'next', iteration for repetition of actions or selection to specify where the flow of control branches in an action sequence.

Causal is the high level, cause-effect relations explaining how and why some part of the world behaves. This information type includes actions and objects in a model with other knowledge to answer 'why', 'actions happen because' of other actions and states which result from actions.

Information analysis considers what information is necessary for the user to carry out the task. This inevitably involves decisions about how much of the task is automated and the design of user-system interaction. Information may be required as input for a user action, as input to an automated computer action, for a human decision, or ultimately instructions may be required on how to carry out the task itself.

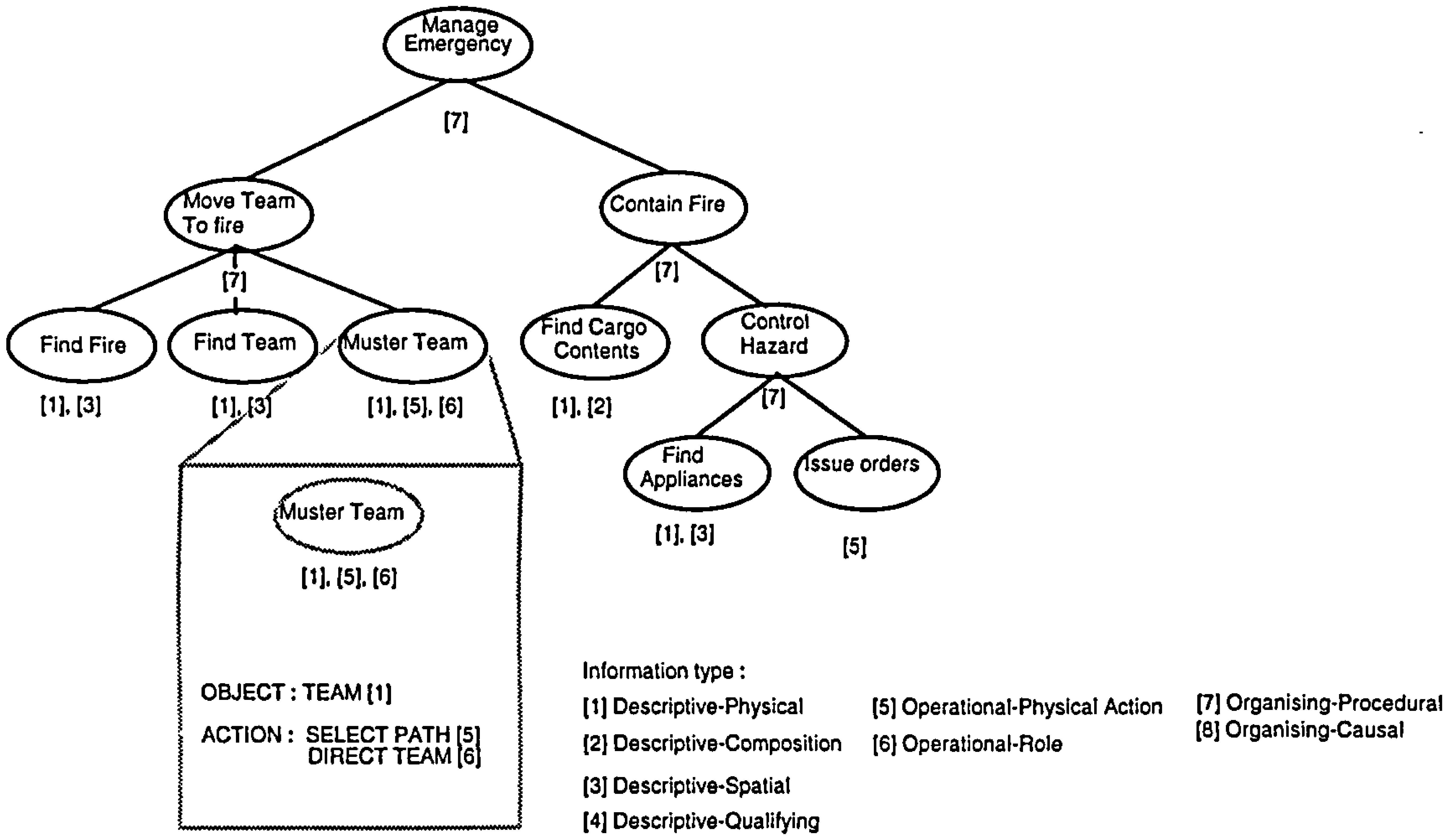


Figure 3.3 Task Model for Emergency Management with details for 'Muster Team' step

If the user requires an aide memoir of how to carry out the task-goal then a summary of a set of actions, i.e. operational-procedural information should be given. In the Muster team sub-goal (see figure 3.3) the captain is given such an aid memoir for the muster team procedure. Most sub-goals required both descriptive-physical and descriptive-spatial information as several procedures are concerned with finding and locating objects (e.g. move compartments, hazards, fire fighting equipment). When the sub-goal requires user action, information on the objects involved is given, and the action specified using operational information, although if the user is familiar with the action, this may not be necessary. However, if users do not know the action/procedure then operational information should be supplied. Information needs have to be considered at the level of action and whole procedures. For procedures, i.e. sub goals realised by several actions, organising information may be required, hence procedural information is attached to sub-goals as one user requirement for the emergency management system was for training. Figure 3.3 shows the task model annotated with information types, together with an inset giving more detail of the action and object required for the Muster team sub-goal.

3.5 Planning the presentation

The planning process is broken down into a number of steps. First media resource(s) must be chosen to satisfy the information requirements of each task sub-goal. These are also validated against a set of cognitive heuristics. Then the presentation techniques available to the designer are analysed. The media resources are then annotated using presentation techniques to generate the Presentational effect for the sub-goal.

3.5.1 Media Selection for Information requirements

The information types are used to select the appropriate media resource(s). Task characteristics influence the modality of media resource used, for instance, verbal media are more appropriate to language based and logical reasoning tasks; whereas visual media are suitable for physical actions involving moving, positioning and orienting objects. If no suitable media resources are available for selection, the designer will have to create additional media as necessary for the information requirements.

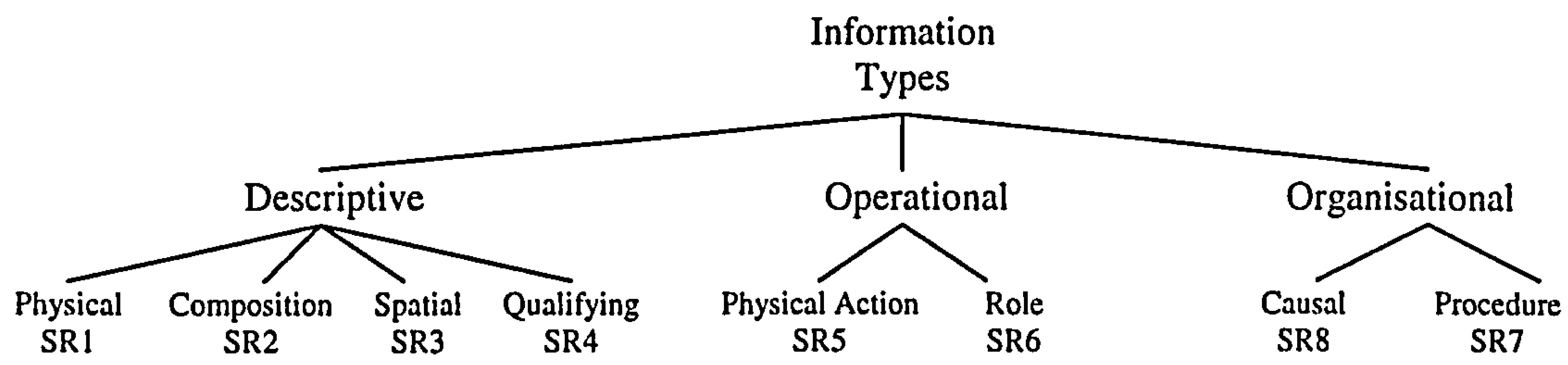


Figure 3.4 Overview of Information types and related Selection Rules (SR)

Each selection rule (SR) links the required information type with appropriate media. An overview of the information types and selection rules is shown in figure 3.4, and in detail below :

Descriptive :

SR1 : if the task sub-goal requires physical information then prefer a visual medium. Language is poor at describing object detail and appearance (see 2.7.1.1 [Bieger & Glock 84]; [Dwyer 67])

Caveat : If an object has to be identified which may be difficult for the user to recognise, use language-text to indicate identity (see 2.7.1.1 [Wadill & McDaniel 92]; & 2.6.1 [Jorg & Hormann 78])

SR2 : if the task sub-goal requires composition information then use a visual medium (see 2.7.1.1 [Bieger & Glock 84]; [Dwyer 67])

Caveat : Language can be used to set the granularity of the components in the image e.g. "Look as sub assembly X." (see 2.6.1 [Hegarty & Just 93])

SR3 : if the task sub-goal requires spatial information then prefer a visual medium (see 2.7.1.4 [Bieger & Glock 84])

Caveat : Language- text captions can be used to identify landmarks and components for spatial information (see 2.7.1.4 [Levie & Lentz 90]; [Dean & Kulhavy 81]; [Ferguson & Hegarty 94])

SR4 : if the task sub-goal requires qualifying information then use a linguistic media (see 2.7.1.1 [Boohrer 75])

Operational :

SR5 : if the task sub-goal requires physical action information then prefer a visual medium (see 2.7.2.2)

(a) animation for complex actions e.g. those with complex paths, or manipulations [Swezey 91]; [Park & Hopkins 93]; [Kaiser, Proffitt, Whelan & Hech 92]
(b) still image for simple actions [Swezey 91].

Caveat : Language should be used to identify and amplify explanation of complex actions, or groups of action; or to qualify actions (see 2.7.2.2 [Boohrer 75])

SR6 : if the task sub-goal requires role information identifying agents or object involved in the action then use a linguistic medium to add this information to the image or animation sequence (see 2.7.2.1)

Organising :

SR7 : if the task sub-goal requires procedural information then prefer a linguistic medium to provide sequencing relations, particularly when the time frame is not constant (see 2.7.3.1 [Palmiter et al 91]; [Burch 73]); animation or image sequences will be useful to support any underlying physical action information.

SR8 : if the task sub-goal requires causal information prefer a linguistic medium to provide key causal relations (see 2.7.3.4 [Mayer & Anderson 91]; [Carpenter & Just 92]; [Park & Hopkins 93]), with animation or images sequences to support any underlying actions being related.

3.5.2 Validation

The selection rules are augmented with a set of validation rules (VR) , which warn about the possible consequences of undesirable combinations of media resources :

Choice of Static 'v' Dynamic media

VR1 Use text or still images for important information which must be attended to as information may be lost from time varying media. Memory for content of dynamic media is generally worse than for static

VR2 Beware that animation media will dominate over static image media, as attention is drawn by default to motion and stimuli which change (see 2.3.4).

VR3 If language to be presented is complex or lengthy, then favour text. If the language is simple or short then favour speech (see 2.5.1 [Deathersage 72])

Re-inforcement

VR4 Present the same message on two (or more) modalities for reinforcement, e.g. an animated demonstration of a procedure is accompanied by voice description. (see 2.6.2)

VR5 Repeating the same content may be useful to summarise or re-inforce, but will be less effective than content which provides more information (see 2.6.2 [Wagenaar 81])

Media Combination & Ordering

VR6 If language is to be combined with animation then favour speech, as reading text will compete with viewing animation (see 2.5.1 [Deathersage 72])

VR7 If different parts of the same message are to be presented in different modalities then cross referencing the message by contact points effects is required; details are given below (see 2.6.1)

VR8 If verbal media are being presented concurrently with visual media, the message on two media should be integrated. Ensure that the language and image share the same topic (see 2.6.2)

VR9 For speech and text, the same wording should be used (see 2.5.1[Grimes 91])

Example

Applying these guidelines to the 'Move Team to Fire' procedure (Figure 3.5), organising-procedural information is required, and is hence assigned to language as the overall instructions for fire fighting are procedural instructions (SR8). As the instructions are likely to be complex, the validation rules favour using text (VR3).

The 'Find Fire' task sub-goal calls on descriptive-physical and descriptive-spatial information, so selection rules (SR1 & 3) favour a still image. To identify the object and its location, language is required (SR1 & 3). The 'Find Team' sub-goal is similar, favouring a still image and language. The validation rules suggest that speech should be favoured for language combined with visuals (VR6). They also warn the designer that the speech may need to be re-inforced with text (VR1) and that the speech, image and text media must share the same subject matter (VR8, VR9).

The 'Muster Team' sub-goal requires descriptive-physical and operational-action information to find the appropriate location for the fire team and then direct the team to the compartment containing the fire. The selection rules favour both language and animation (SR 5) for complex operational information which has abstract instructions and physical information about the path to be taken. Figure 3.4 illustrates the media resources chosen for each sub-goal.

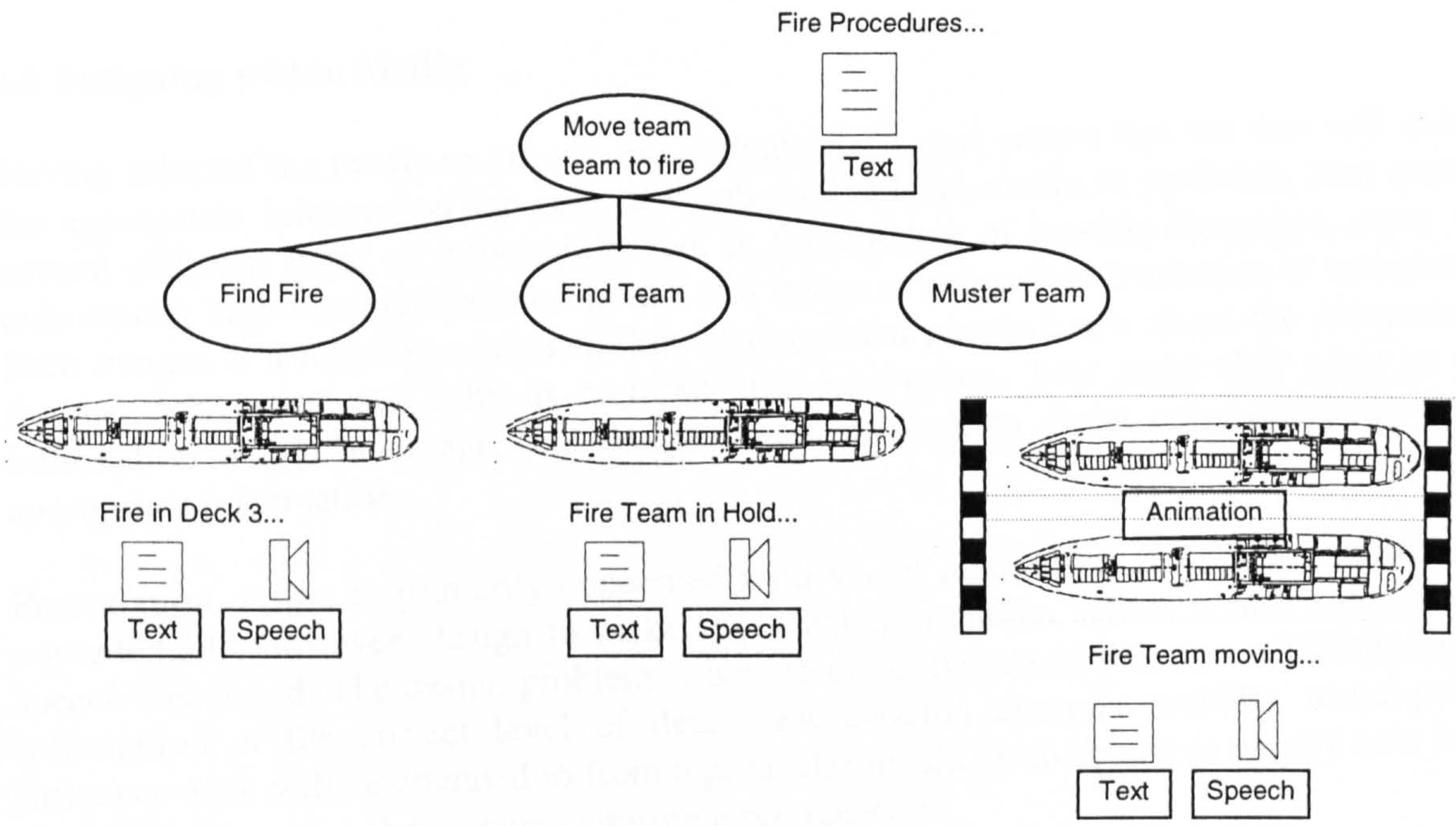


Figure 3.5 Example of Media Selection for 'Move Team to Fire' procedure. The 'find fire' and 'find team' steps are given an image, speech and text; the 'muster team' step is given an animation, text and speech.

3.6 Designing within Media

Having selected the media resources, the designer must now ensure that the user will extract the appropriate information for each task sub-goal. Visual media in particular may contain several different types of information and, in the absence of specific directions, users will only extract high-level information from a still image or animation. Extraction of information from images is a function of domain knowledge- what people know about the components, the task which indicates what is required, user motivation- how much they want to find information within an image, and design of the medium to direct users' attention to the appropriate information.

Presentation design is primarily concerned with visual medium as the reading sequence is unpredictable; however, design to make important information salient is also necessary for speech and sound. The design problem is how to direct the user's attention to the appropriate information at the correct level of detail. Presentation analysis provides techniques to influence what will be attended to from a particular media resource and to specify how long a piece of information should persist during a presentation.

Three sets of guidelines are given which are applied to the information types and media selected. These are Image Presentation (IP), Animation Presentation (AP) and Language Presentation (LP) guidelines. Table 5.2 gives an overview of how these guidelines address each information type and media; the following sections then provide more detail :

	Image Presentation (IP)	Animation Presentation (AP)	Language Presentation (LP)
Physical	IP1-IP2		
Composition	IP3-IP4	AP1-AP6	
Spatial	IP5-IP6		
Role		AP7-AP8	
Physical Action	IP7-IP8	AP9-AP11	
Procedure		AP12-AP13	LP1-LP2
Causal			LP3-LP4

Table 5.2 Overview of Designing within Media

3.6.1 Still Image

3.6.1.1 Physical Information :

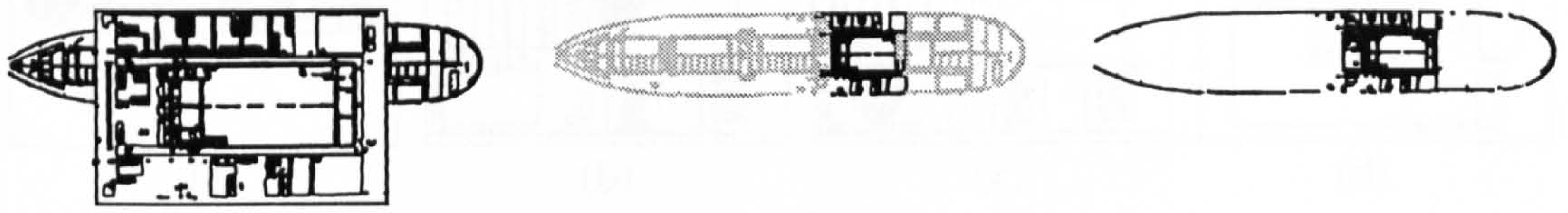
IP1 : If an unusual or complex object is required, then it is important to show the whole object, unobscured, as identification will initially use outline (see 2.7.1.2 [Biederman & Ju 88]). E.g. if in the find team sub-goal a specific deck is to be identified, the image should be shown whole to aid identification.

IP2: If the task requires a specific object to be identified, or requires details of object properties, then use colour and texture in images to identify the object (see 2.7.1.2 [Price & Humphreys 89]).

3.6.1.2 Composition Information

IP3 : Information will initially be extracted from the image at scene level- identity of major objects with very little descriptive detail (see 2.7.1.3 [Pezdek 88];[Friedman 79]; [Mandler & Ritchey 77]). Apply an attentional effect (e.g. highlight) if an object is important (see below).

IP4 : By default objects focused on will be those which are : bright in colour, set apart from other objects, larger in size, shown in more detail, in sharp focus, or nearer the front of the scene (see 2.3.3). These should be objects which are important in the task. E.g. if a part of the deck is important then enlarge it (see figure 3.6.1)



(a) (b) (c)

Figure 3.6.1 Examples of using (a) size, (b) sharp focus and (c) detail to set focus on a particular part of the ship's deck

3.6.1.2 Spatial Information :

IP5 : If a particular object needs to be located accurately, then set it as a landmark (see 2.7.1.5 [Macnamara 91]). Landmarks should be perceptually salient objects, or objects with which the user is already familiar and knows the location of. These will be located more accurately and will help organise the location of other objects. E.g. the objects which are in focus in figure 3.6.2 would be accurately located as a landmark.

IP6 : If several objects are to be located as landmarks, it is important to divide the image into sub-areas if possible and then select a landmark for each area (see 2.7.1.5 [Macnamara 91]).

3.6.1.3 Physical Action Information

IP7 : For physical action information given in as a sequence of still images it is important that breakpoints are shown e.g. start and end points of the action, and any changes in direction or speed are shown within the image sequence for the action to be identified (see 2.7.2.3 [Newtson & Enquist 76]). E.g. if animation was shown of the fire team following a complex path, it would be important to show the points at which the team changed direction.

IP8 : Pictorial symbols may aid in the identification of physical action information, particularly in still images (see 2.7.23 [Carello, Rosenblum & Grosofsky 86]; [Friedman & Stevenson 80]). These should be chosen to make the style or path of motion more important. E.g. an arrow symbol can be used to show the path of fire team through the deck in the muster team sub-goal.

3.6.2 Moving Image

3.6.2.1 Composition Information :

AP1 : To start a new sequence and provide context, use an establishing shot : show the whole of scene (see 2.7.1.3). E.g. to introduce the deck at the start of the fight fire sub-goal, an establishing shot would be used to show the whole deck for context (see figure 3.6.2a)

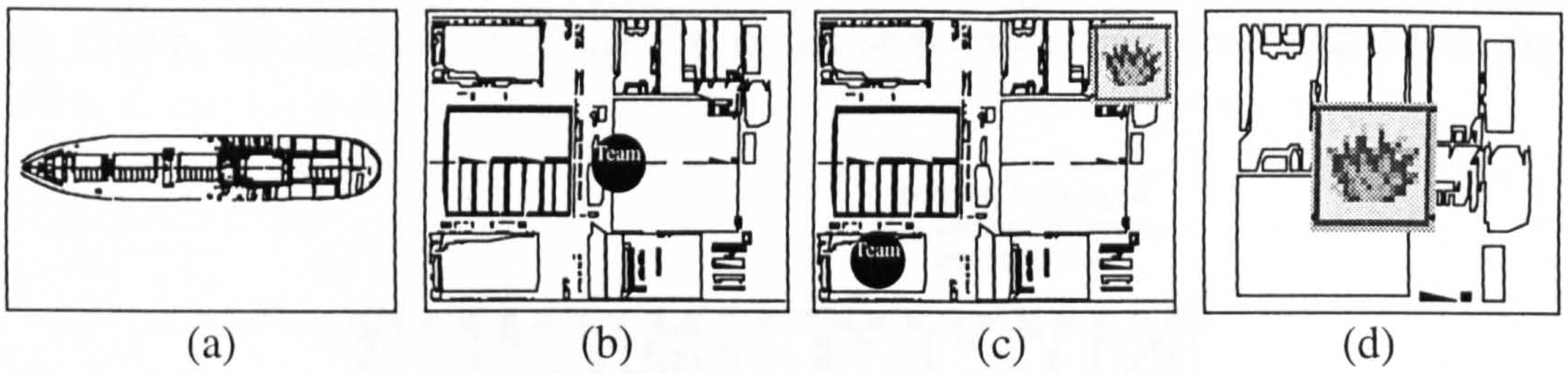


Figure 3.6.2 Use of (a) establishing shot, (b) full shot, (c) tight two shot and (d) close up to show fire team and fire.

AP2 : To provide detail of newly introduced object, or context for an object use a full shot; show the whole object in the frame with surroundings scene (see 2.7.1.3). E.g. to introduce the fire team at the start of the find team sub-goal, show it centred within the scene, together with its immediate surroundings (see figure 3.6.2b)

AP3 : To imply a relationship or compare two objects use a tight two-shot; show the two objects together in the same frame (see 2.7.1.3). E.g. in the muster team to fire sub-goal, to show the relation between the fire team and fire use a tight two-shot (see figure 3.6.2c)

AP4 : To make an object very salient use a closeup; show the object filling the frame completely (see 2.7.1.3). E.g. to focus on the fire, use a close-up (see figure 3.6.2d)

AP5 : A zoom effect, where the image is shown expanding in the frame, can be used to provide context in the change from one type of shot to another if the image is complex or unusual (see 2.7.1.3 [Marcel & Barnard 79]). E.g. In the find fire sub-goal, in changing from an establishing shot of the deck to a full shot of the fire, a zoom effect should be used as the image is complex.

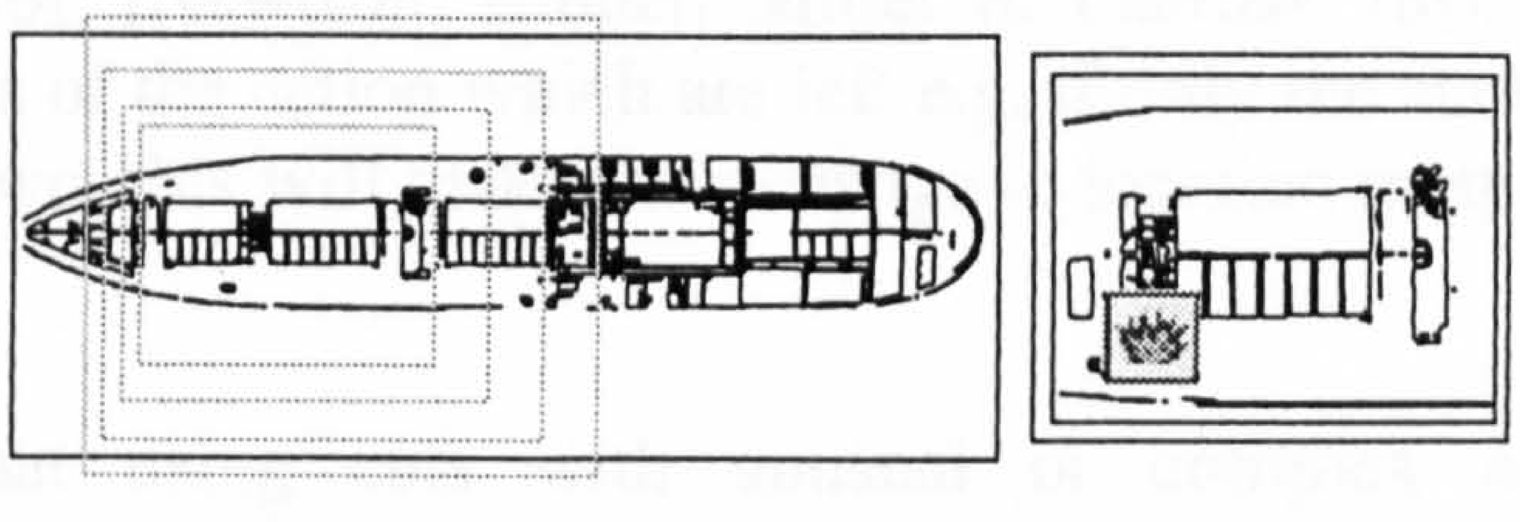


Figure 3.6.3 Use a zoom to change from an establishing shot to a full shot as the deck image is complex

AP6 : If the composition information is complex or unusual, show the whole of the zoom effect, not just the start and end views (see 2.7.1.3 [Salomon 79]). E.g. in figure 3.6.3, as the ship is complex, it is important to show the image expanded out, rather than switching to show the full shot.

3.6.2.2 Role Information :

AP7 : When displaying motion, keep the ground over which an agent moves still to allow the agent to be identified; avoid other unnecessary motion in the background (see 2.7.2.5 [Hochberg 86]; [Kraft 87]). E.g. in the muster team sub-goal, motion of the fire team is being shown, the appearance of the fire in the ship should not be changed (see figure 3.6.4)

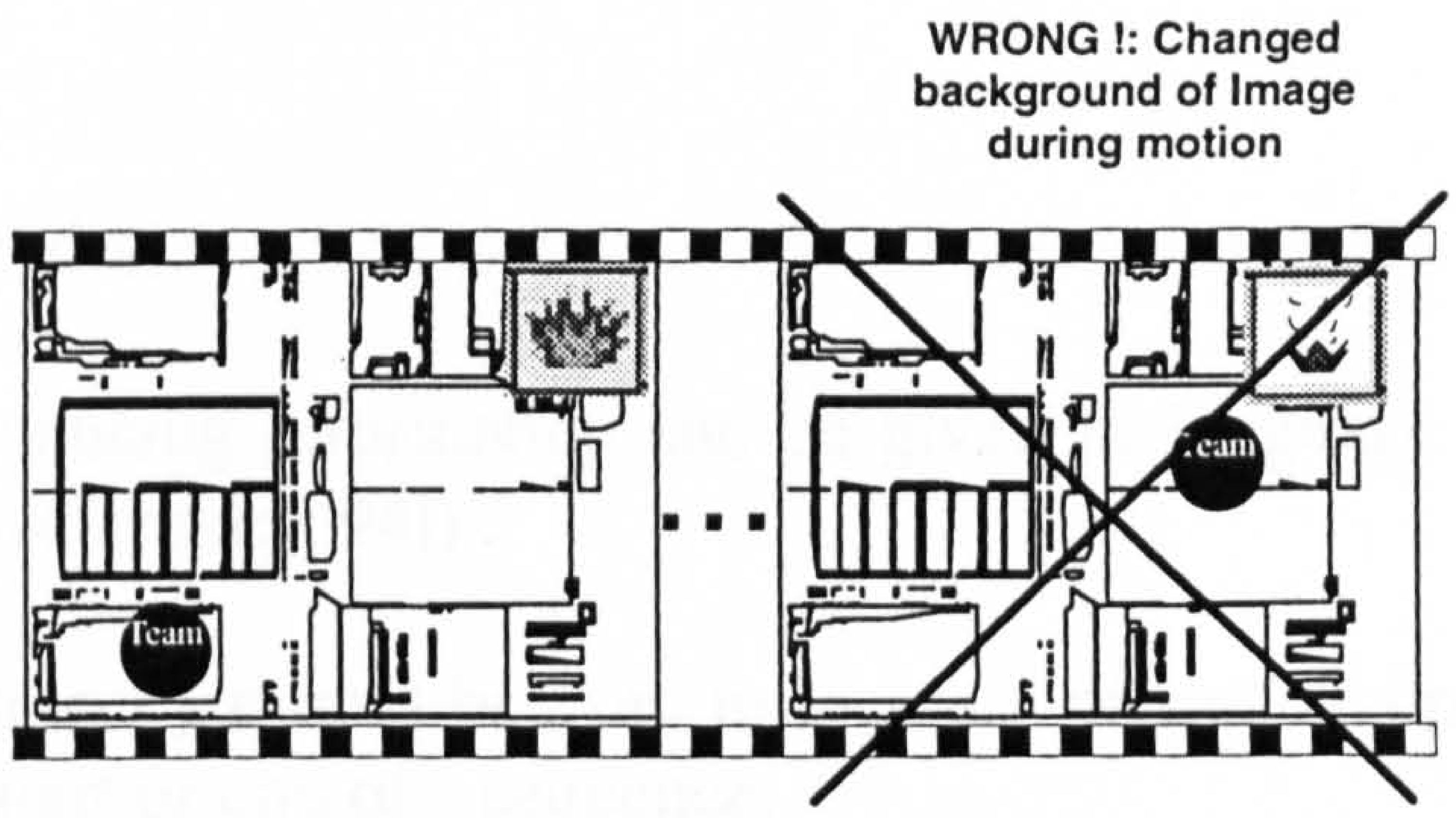


Figure 3.6.4 Avoid changing the image background (e.g. appearance of fire) whilst showing the motion of the fire team across the deck

AP8 : If the agent is changing location, ensure that it has a constant appearance during its motion to help preserve identity (see 2.7.2.5 [Hochberg 86]; [Kraft 87]). E.g. in muster team, the fire team's visual appearance should be maintained across the motion to preserve identity

3.6.2.3 Physical Action Information

AP9 : Recall of the details of action may also be improved if it is shown in slow motion (see 2.7.2.4 [Newtson & Ridner 79]). E.g. if the motion of the fire team is particularly complex in muster team, then it should be shown in slow motion.

AP10 : If part of an action is unimportant, then a cut may be used to remove it (see 2.7.2.4 [Hochberg 86], [Newtson, Rinder, Miller & Lacross 78]). This will focus the viewer upon the parts of the action which are left e.g. if only the start and end position of the fire team are shown, this will make their change in location more important than their path.

AP11 : Beware that using cuts with unusual or complex actions may disrupt comprehension (see 2.7.2.4 [Newtson, Rinder, Miller & Lacross 78]). E.g. if the motion of the fire team is complex in the muster team step, then a cut may confuse.

3.6.2.4 Procedure Information :

AP12 : To structure information in animation sequences use either a wipe / dissolve or a cut (see 2.7.3.3 [Kraft 86]; [Burch 73]). These emphasise a section boundary in a sequence. The sequence will be assumed to be continuous unless a cut, wipe or dissolve is encountered.

AP13 : If a large change in sequence is required, such as a flash forward or back, then a wipe or dissolve should be used as an explicit signal (see 2.7.3.3 [Madsen 78]). Cuts in general signal a smaller change in time. E.g. a cut could be used within the muster team animation, as it would only be a small change within the sub-goal; whereas a wipe / dissolve should be used to show the larger change to the next major sub-goal.

3.6.3 Linguistic Media

3.6.3.1 Procedure Information :

LP1 : Procedure sequencing information can be given in language by cue phrases (see 2.7.3.2 [Betsgen & Costermans 94]) :

- A cue (e.g. 'at time x') should be used to locate a particular important time point in the task, such as the start or end of a sequence.

- Cues can be used to order the sequence ('next', 'then').

LP2 : Structure text to indicate a procedures order by formatting into lists or frames (see 2.7.3.2 [Armbuster, Anderson & Mayer 91]). E.g. in providing the plan for the fight fire goal, the text should be formatted into sections for each sub-goal (see figure 3.6.5)

Fight fire emergency procedure

i) Locate available fire team

ii) Search ship for the location of the fire

iii) Muster team to fire's location

Figure 3.6.5 Use lists and advanced organisers to indicate order and importance in complex procedures

3.6.3.2 Causal Information :

LP3 : To indicate causal relationships use cue words: e.g. 'because, in order to, consequently, resulting-in' (see 2.7.4.2 [Kintsch & Yarbrough 82]; [Meyer, Brandt & Bluth 80])

LP4 : Advanced organisers can be used to define a complex sequence or causal relationship. Use titles to support topics (see 2.7.4.2 . [Ausubel 62])

3.7 Attentional Design & Attentional Bar Chart

Presentation techniques are used to help direct the user's attention to important information and specify the desired reading order. The design problem is therefore :

i) To plan the attentional thread to guide the user's viewing and reading sequence

ii) To design effects and combinations of media to make important facts salient and reinforce the contact points between different media.

iii) To determine timing and synchronisation so that the user has sufficient time to assimilate the content.

Attentional bar charts are used to plan the sequence and duration of media delivery. First the information types are ordered in a first cut sequence. The selected media are added to the bar chart to show which media stream will be played over time. The information model defines the high level presentation order. Decisions on timing depend on the content of the media resource, e.g. length of a video clip, frame display rate. The display duration of text and images require analysis of the content being presented and the use of user tests where necessary to decide the appropriate display time. Presentation timing of dynamic media, such as in sound recording or animation, should be under user control with replay facilities.

Focus control actions are added to the first cut presentation script to either make specific information within a medium more salient or to draw the user's attention to the message thread between media.

The attentional bar chart for the find fire subgoal is shown in Figure 3.7. The task sub-goals are given along the top of the chart. This represents the presentation of the task over time.

In the first column of the chart is the information types which are to be presented. In the second column is the media and presentation techniques which are to be used to present the content. The columns which follow show which media and technique are to support each task step. A black line shows that a particular medium / technique is to be in focus, an arrow show that focus is to be shifted to a new media / technique, while a grey line shows the medium / technique is no longer in focus but is still available.

In figure 3.7, focus is initially set upon the text resource to give information on the procedure. It then shifts to speech, giving descriptive information as to the fire's location, and to an image of the deck. The text resource for the procedure is also left on view, so it is shown on the graph as being in partial focus.

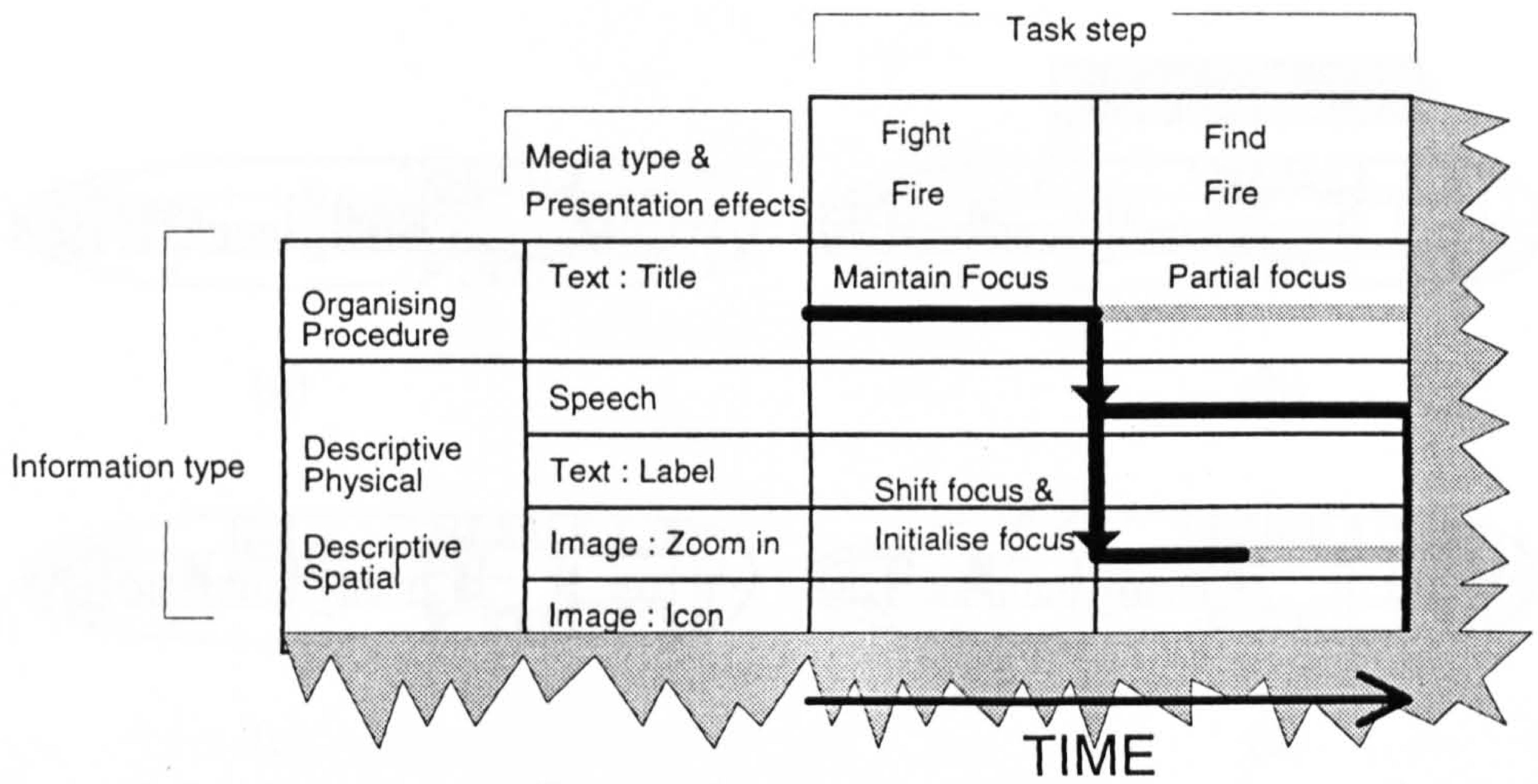


Figure 3.7 Fragment of attention bar chart for 'find fire'sub goal

The need for focus shifts between each information component are identified and attention marker techniques are selected to implement the desired effect using guidelines. The guidelines are again based on the media selected : Image Attention (IA), Animation Attention (AA), Text Attention (TA) and Speech / Sound Attention (SA). Table 5.3 gives an overview of the guidelines.

Image Attention (IA)	Animation Attention (AA)	Speech & Sound Attention (SA)
IA1-IA7	AA1-AA7	SA1-SA4

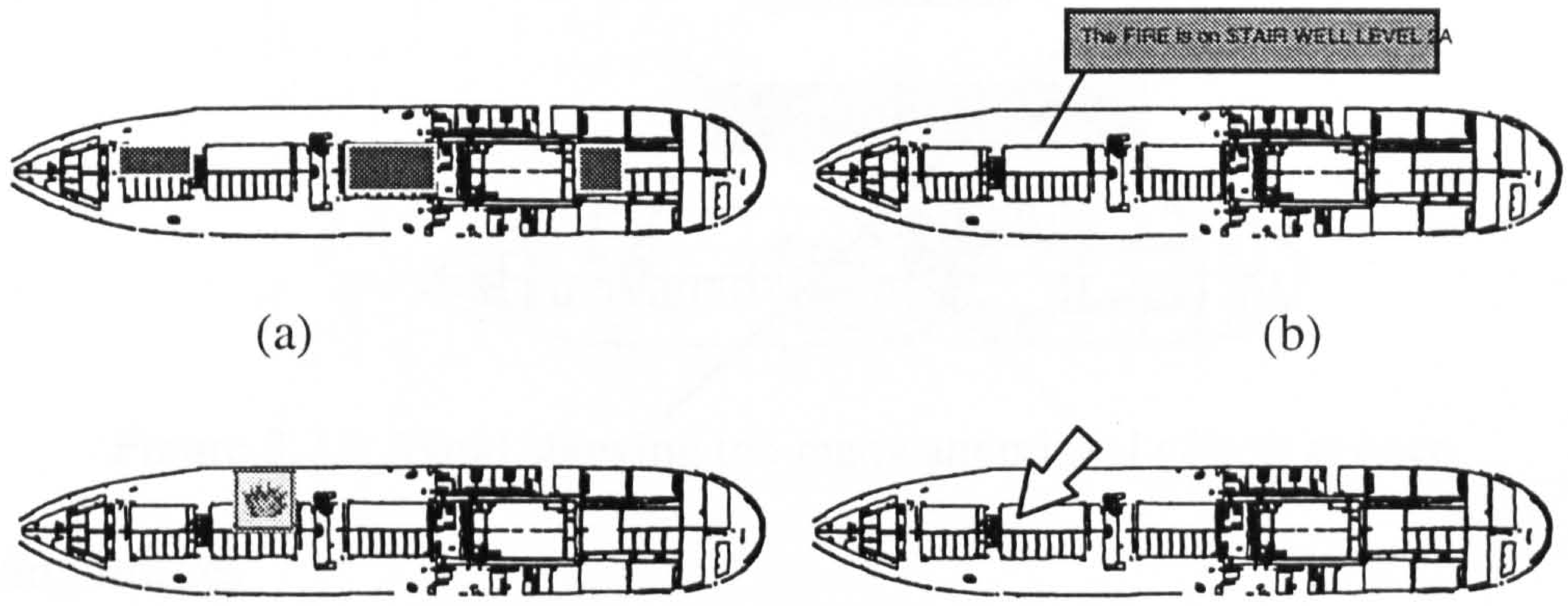
Table 5.3 Overview of Attentional Guidelines

3.7.1 Still Images

IA1 : Use a highlighting technique (colour, shape/size change) to draw attention to important objects (see 2.3.3 [Treismann 88]; [Arnheim 68]). By default only the scene level will be focused upon unless the user is motivated for closer inspection; or an object in the scene is unusual.

IA2 : To draw attention to a group of spatially distributed objects set a common visual attribute, e.g. change to the same colour (see 2.3.3 [Treismann 88]; [Arnheim 68]). To emphasise a group of co-located objects highlight the background or draw a box around the objects (figure 3.7.1a). The highlighted area will set the granularity of the user's attention.

IA3 : Icons may be placed by or on an object to draw attention (see 2.3.3 [Fleming 87]; [Kennedy 82]). Use if highlighting may obscure detail (figure 3.7.1c)



(a) (b)

(c) (d)

Figure 3.7.1 Attentional effects to focus on a particular location on the deck.

IA4: An arrow may be used as a pointer to focus on an object (see 2.3.3), useful to locate small detailed components (figure 3.7.1d)

IA5: Labels linked to objects in an image are useful for drawing attention and providing supplementary information (see 2.6.1) e.g. identity of part of the deck in figure 3.7.1(b). Dynamic revealing of captions is particularly effective and can be used to direct the user's reading sequence.

IA6 : Avoid showing an object in motion or using a highlighting technique when the user is extracting information from an image; allow at least a second before changing the image (see 2.3.4; & 2.3.3 [Hochberg & Brooks 78]). E.g. in the muster team sub-goal, the attentional effect icon should not be shown during the motion of the fire team; it should be revealed prior to the motion, or the motion is freeze framed (figure 3.7.2)

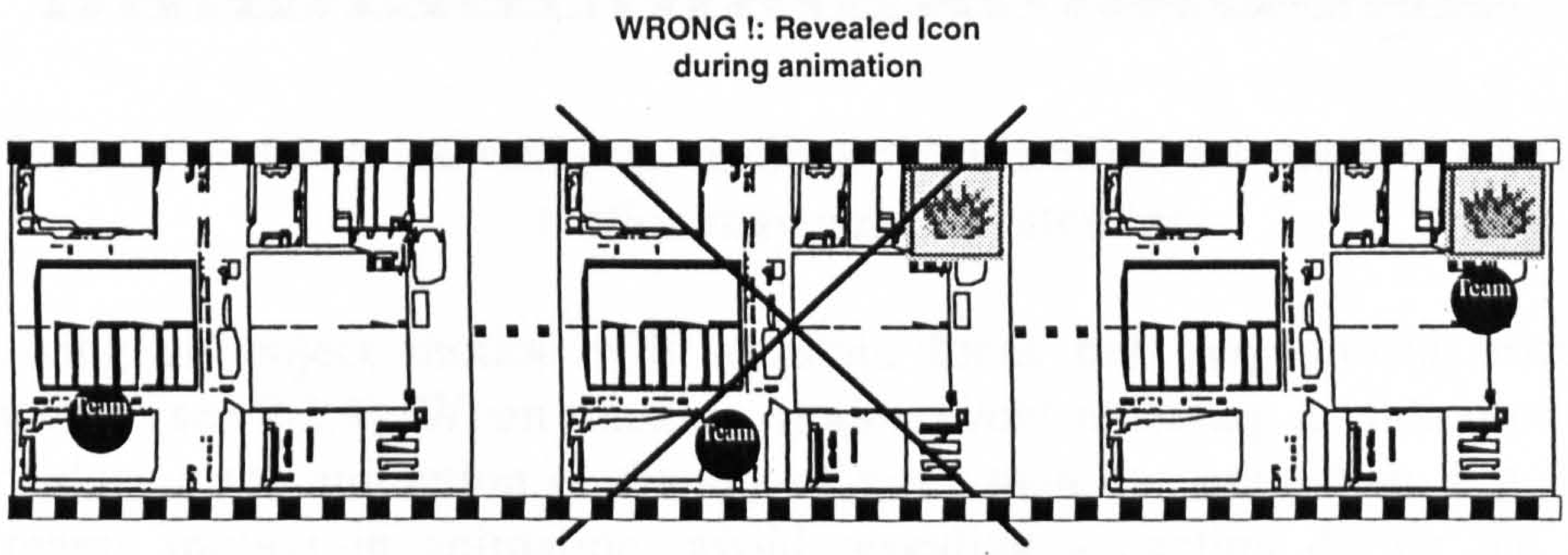


Figure 3.7.2 Avoid using an attentional effect (e.g. revealing an icon) during motion of the fire team

IA7: Beware of using too many highlighting techniques within an image at once; instead sequence highlights to move attention from one object to another (see figure 3.7.3).

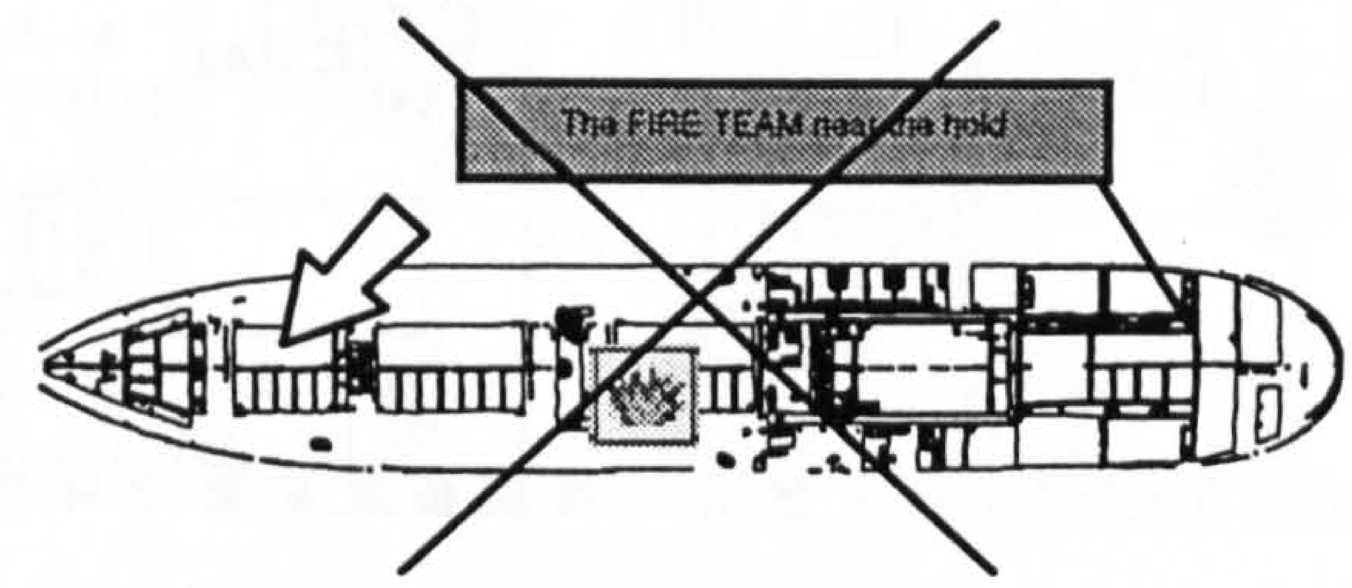


Figure 3.7.3 Avoid showing too many attentional effects at once

3.7.2 Moving Images

AA1 : Onset of motion will attract attention; but it will also shift focus away from static objects in the scene (see 2.3.4).

AA2 : If attention is already focused (such as reading a text or label, or attending to another animation) motion onset may not automatically shift attention (see 2.3.4 [Tepin & Dark 92]). E.g. If a caption is shown for muster team, focus will be maintained on it until it is read, missing the start of the animation (see figure 3.7.4).

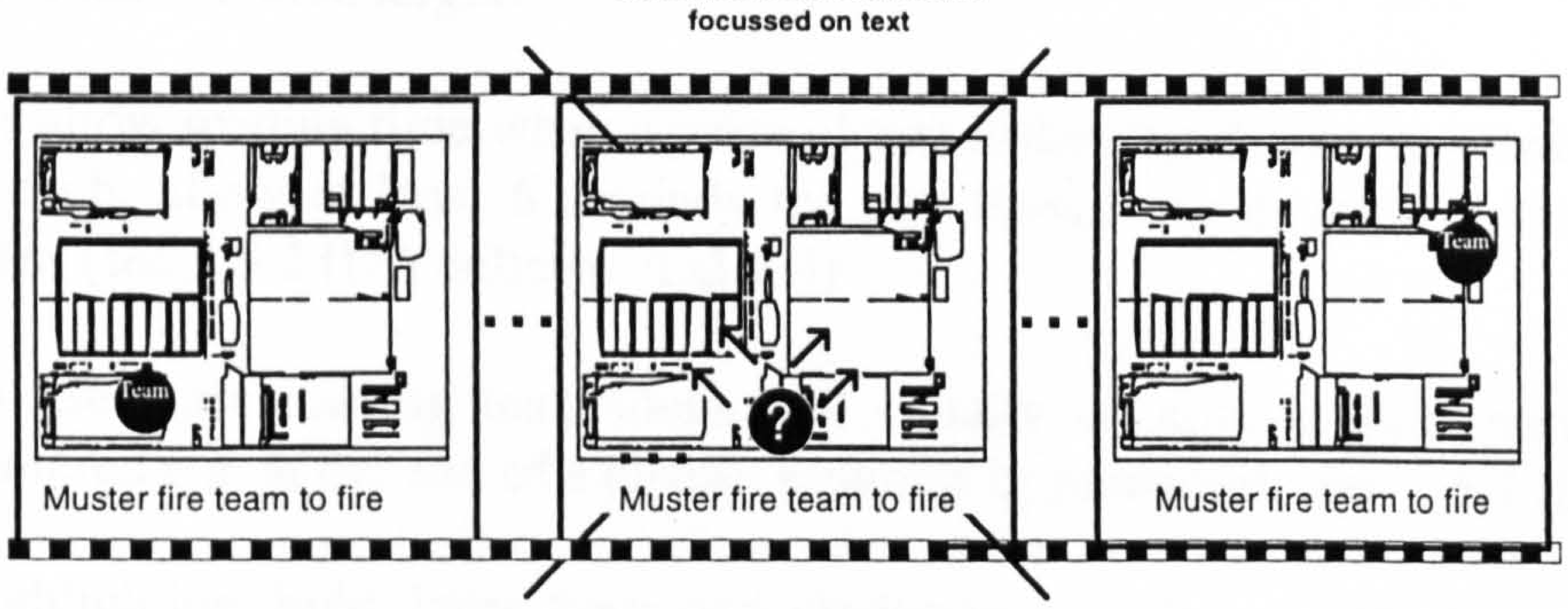


Figure 3.7.4 If attention is focused (e.g. reading a text caption) the onset of the fire teams motion may not gain attention

AA3 : Tracking an object motion will maintain focus but may prevent focus shifting elsewhere (see 2.3.4). When tracking motion avoid revealing a caption or attentional effect during the animation or display of more than one animation. E.g. In showing the teams motion in animation, avoid revealing a caption during the animation; attention may not shift because it will be focused on the animation

AA4 : If an object appears or moves across the ground of an image, it will be more attention grabbing than internal movement of an object's components (see 2.3.4 [Hillstrom & Yantis 94]). E.g. If an animation shows the motion of the fire team across the deck, as well as internal motion of the fire changing state as it is extinguished; attention is likely to shift and maintain focus on the motion of the team, rather than the state change of the fire (see figure 3.7.6)

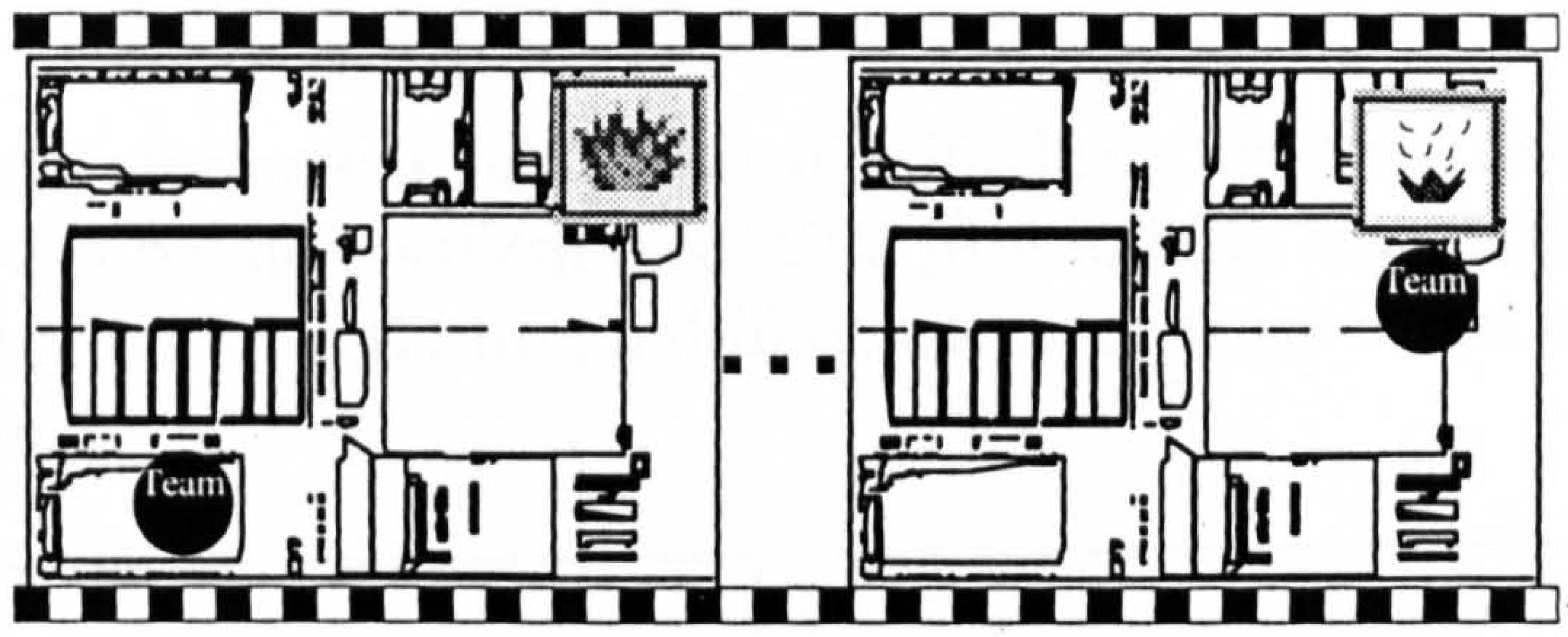


Figure 3.7.6 If an animation shows both state change and position change, focus will usually be given to position change e.g. the change in the position of the fire team will be favoured over the change of state of the fire.

AA5 : Cuts, wipes and dissolve effects in animation will re-set attention (see 2.3.4 [Reeves 93]). Allow the user time to focus on the new image after a wipe or dissolve.

3.7.3 Text

TA1 : Generally an image will be focused on before a text (see 2.5.2 [Brandt 54]). If focus is required on the text prior to the image, the text should be displayed before the image or make the text area larger.

TA2 : Always allow reading time when a piece of text is displayed. Simple words require 200 msec each, allow at least 6 seconds for two lines of text, more if the sentence is complex (see 2.5.2 [D'Ydellawe et al 93])

TA3 : When users are reading text, focus will usually be able to shift until a break is encountered e.g. at the end of a clause, sentence or paragraph (see 2.5.2 [Duffy 93]).

TA4 : Use highlighting, bold, large fonts or underlining to make a particular word or clause of text stand out (see 2.3.2 [Lewis & Walker 89]; [Glyn, Britton & Tillman 85]).

TA5 : To direct attention to the required part of a text use paragraphs and titles as entry points; bullets and lists to guide attention (see 2.3.2)

3.7.4 Speech & Sound

SA1: Use speech and sound to alert users to important information (see 2.3 [Posner 76]). Sound will focus attention initially over visual information; but attention may shift to visual media after a short time.

SA2 : Only a single strand of speech or sound can be in focus at once. Speech will usually gain focus over sound (see 2.2.1 [Patterson 82])

SA3 : Multiple strands of speech or sound will interfere with each other and distract focus (see 2.2.1)

SA4 : To emphasise information in speech use loudness or rate; louder or more slowly spoken words will be more salient (see 2.2.2 [Rubin 80]). E.g. in figure 3.7.7 providing the goal information for fight fire, a different tone of voice could be used to separate it from the sub-goal.

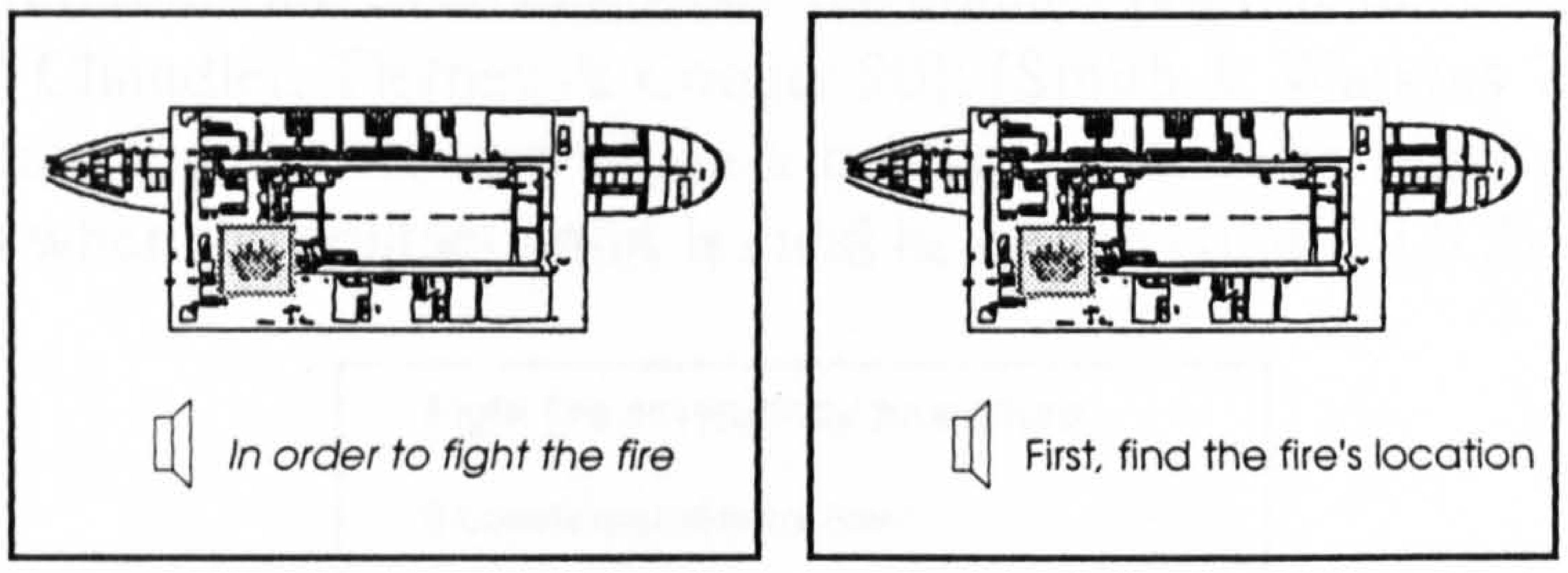


Figure 3.7.7 Use different tone or voices to separate task steps.

3.8 Contact points between media

When combining verbal media and visual media special care must be taken to ensure that the message thread can be followed between the two media. 'Contact Points' are places in the presentation where the verbal part of the presentation needs to be related with the visuals eg for details of an object's appearance, or how to perform an action (see 2.6). The problems are

i) how to provide linking references between language and visuals. A contact point reference can either be 'direct' in which the language explicitly references the visuals eg 'look at X'; or indirect, in which the whole visual is referenced eg 'see figure below'.

ii) how to ensure the message thread can be followed between visual and verbal media.

Contact points are added to the presentation bar charts to show linking references and synchronisation between the media streams. Contact points use the same highlighting and attention directing techniques as described above. The following set of contact point guidelines (CP) are applied to each contact point :

CP1 : Use a direct contact point if the connection between information in an image and language is important (see 2.6 [Levie & Lentz 90]; [Willows 79]) e.g. direct the user's attention to the object in the image by highlighting the object which is being spoken about "look at the cargo hold, [highlight cargo hold]", or reveal a text caption and an arrow pointing to the object (see figure 3.8.1)

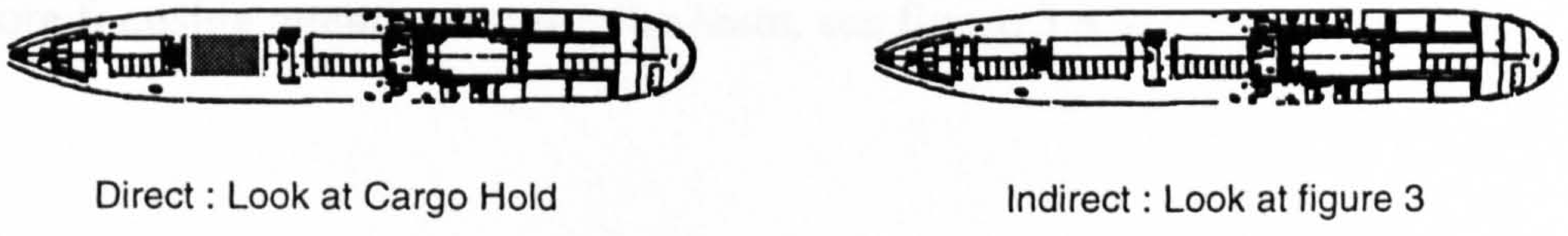


Figure 3.8.1 Use of direct 'v' indirect referents to a particular object or the whole image

CP2: Use indirect contact if the connection between information in an image and language is less important - direct attention to the media resource (see 2.6.1) , e.g. 'look at figure 1' (see figure 3.8.1)

CP3 : Ensure that referents are available when the contact point is made, and are in focus (see 2.6 [Sweller, Chandler, Tierney & Cooper 90]; [Smith & Watkins 72]). Reveal text and image elements together if they share a contact point. Use highlighting techniques in visual media when the contact point is cued in speech (figure 3.8.2)

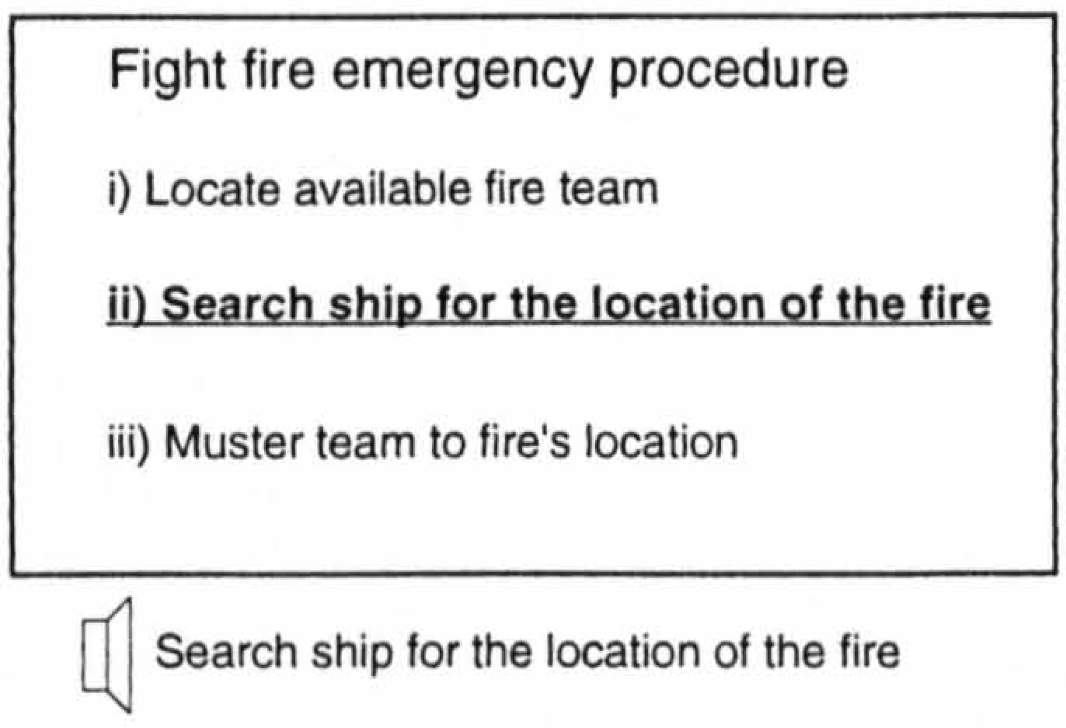


Figure 3.8.2 Highlight the speech contact point in text for the fire procedures

CP4 : If language is being used to set context, it is important that the correct level of identification is set, describe the whole scene; however, if particular objects are important, then identify them with a caption (see 2.6.1 [Jorg & Homann 78]). Thus in the find team sub-goal, the deck should be described first as whole, then specific parts (figure 3.8.3)

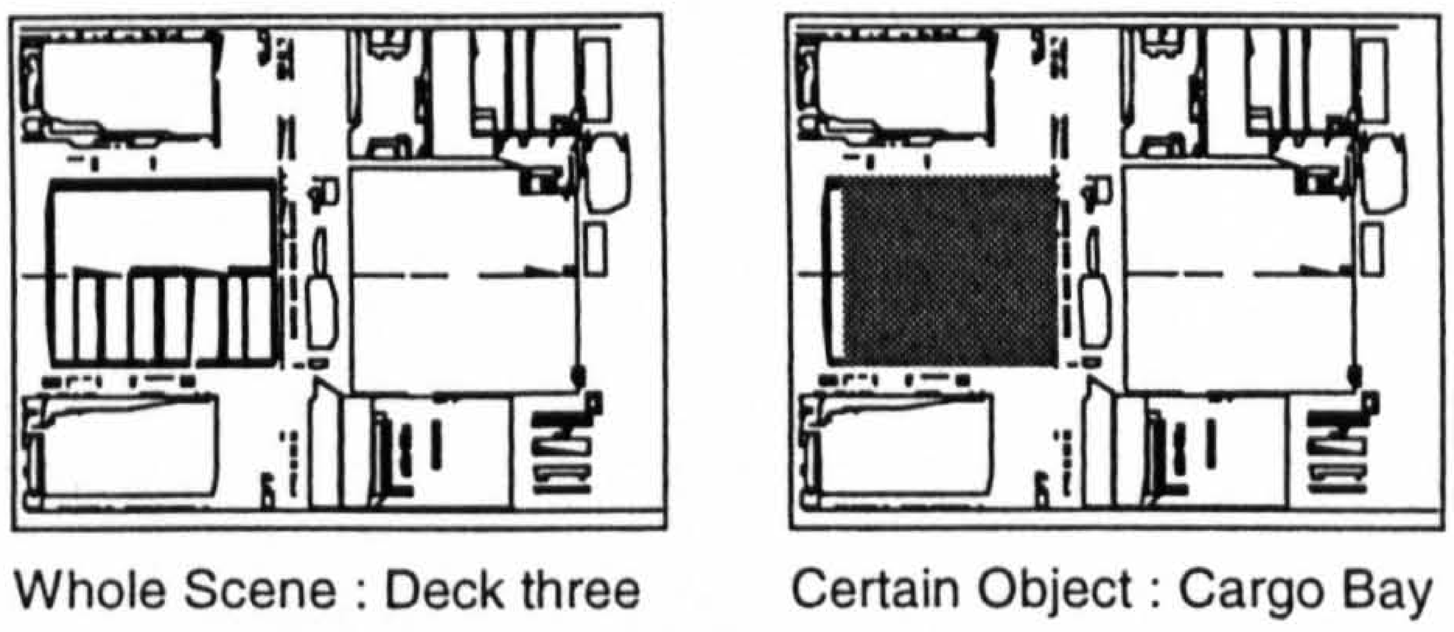


Figure 3.8.3 Identify the scene as a whole first, then identify objects with it.

CP5 When language is combined with visual media, language should be presented before visual media to 'set the scene' and direct the user's attention to information within the image (see 2.6.2 [Baggett 84]).

CP6 : Allow time for contact point to be formed e.g. pace the presentation to allow inspection of image (see 2.6.1), or speech. E.g. in fight fire, allow time for the image to viewed, before focusing attention on the fire team, see figure 3.8.4.

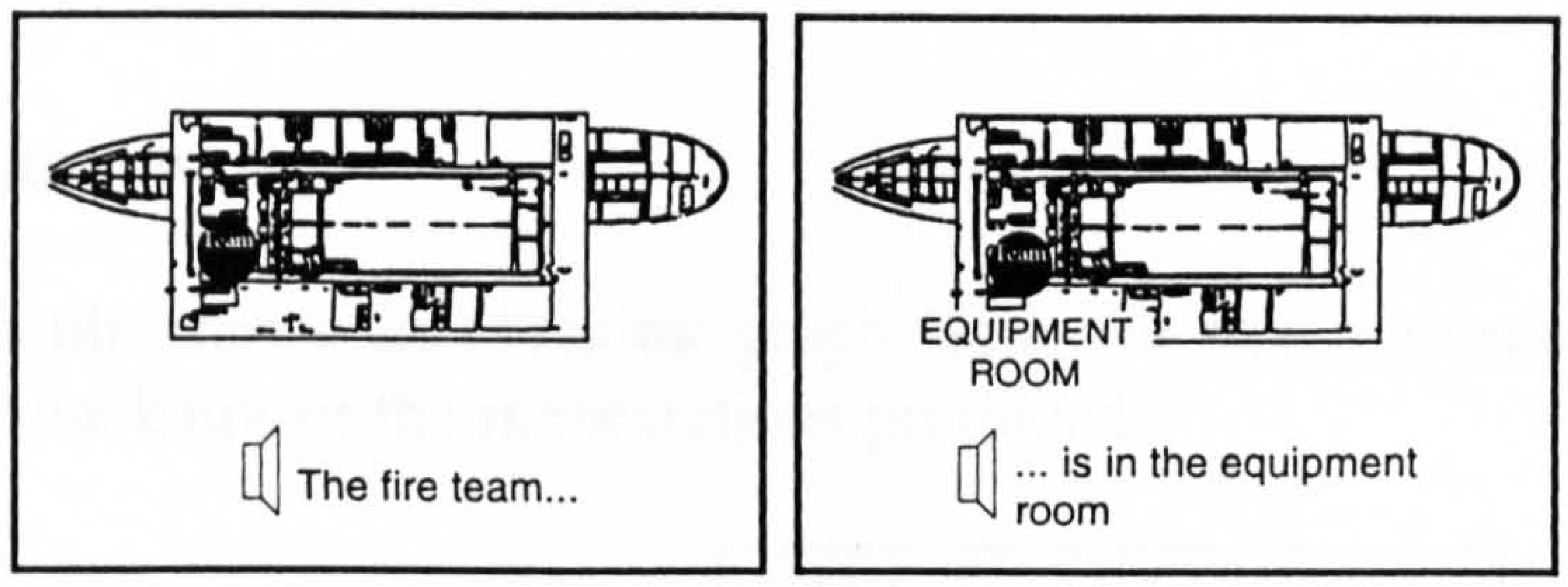


Figure 3.8.4 Pace speech to allow time for the image and labels to be viewed.

3.9 Case Study Example

The following example shows an attention graph from the first procedure 'Move Team to Fire', together with mock ups of the presentations produced.

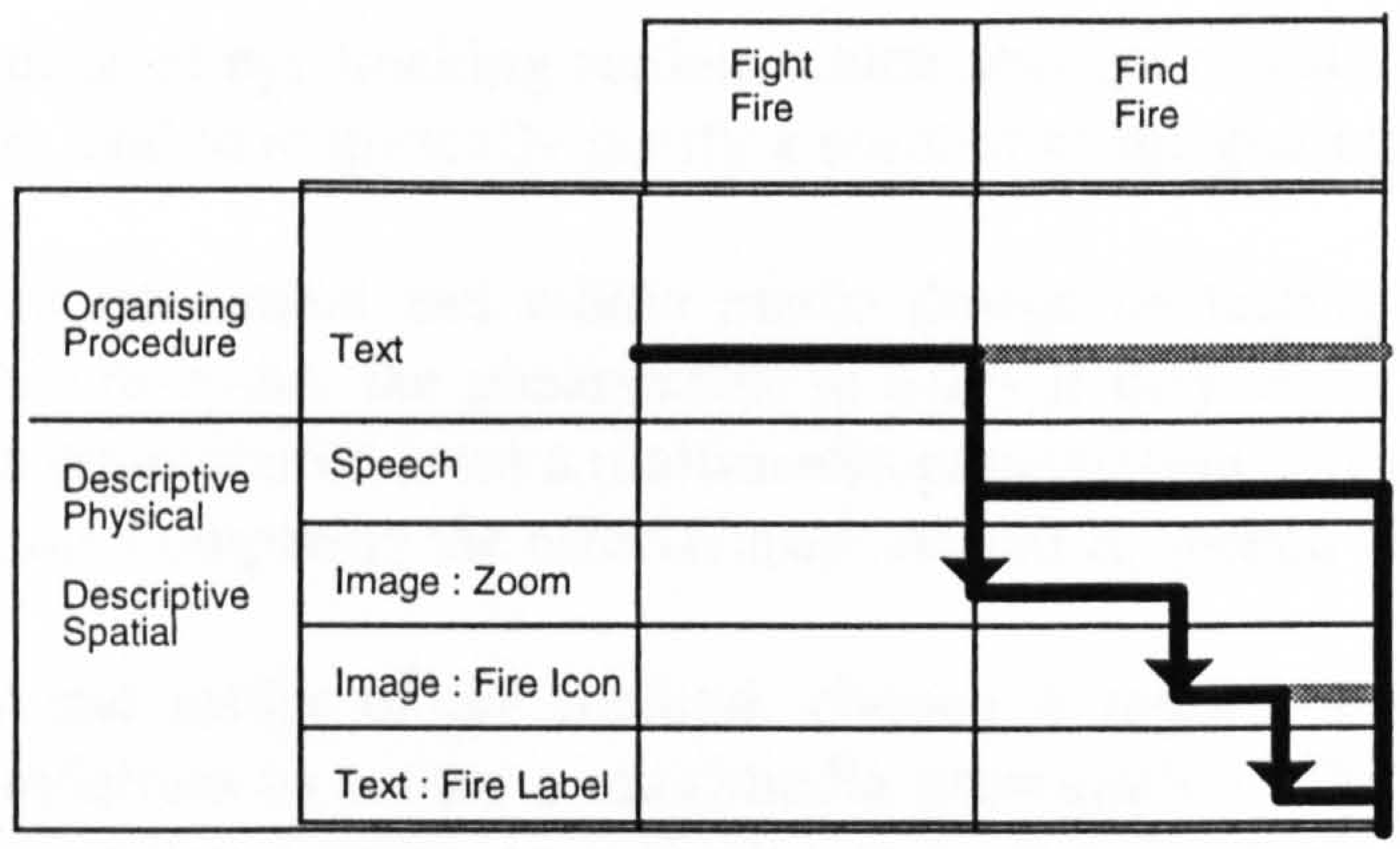


Figure 3.9.1 Attention Bar chart for initial step of fight fire task

A text resource is designed using a title and a numbered sequence. The title will emphasise the importance of the text, it is left on view to act as an organiser for the task steps (LP2).

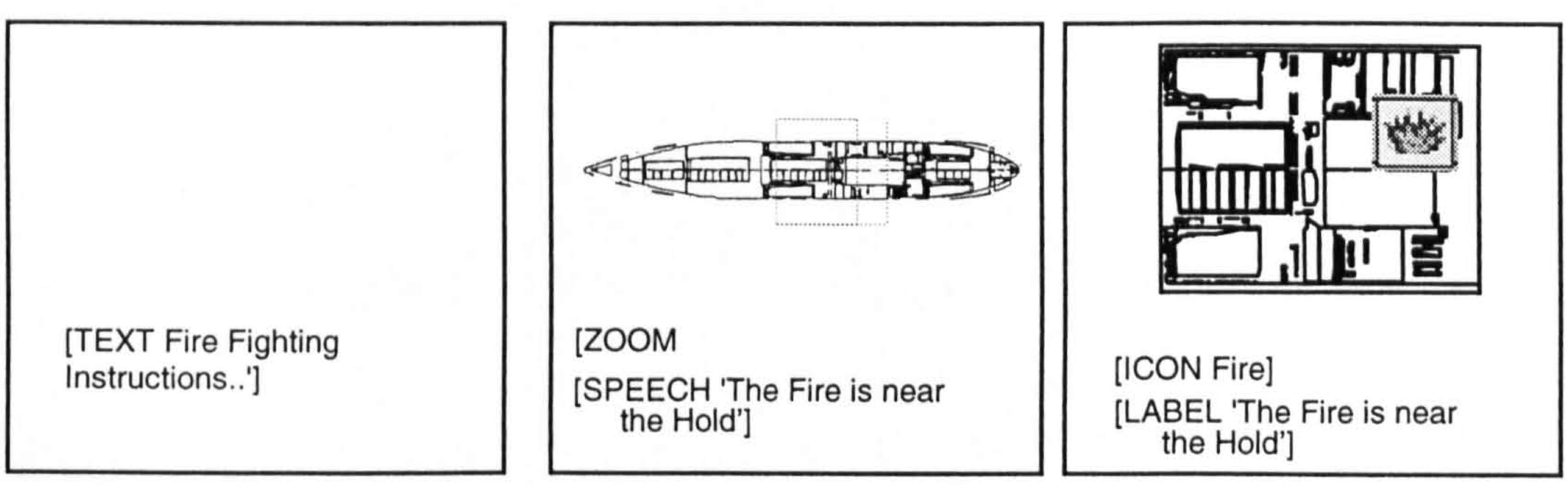


Figure 3.9.2 Presentation Effects for 'find fire' sub-goal

To locate and identify the fire's location an image is used. This is initially shown whole in outline to allow the deck to be identified (IP1). A zoom effect is then used to change from an establishing shot of the deck (AP1) to a full shot of the fire (AP2). This will provide context concerning the fire's location. Speech is employed to provide details of the location of the deck and of the fire within it. As the zoom effect will hold focus on the image. The speech is presented as the image is shown, to set the scene (CP5).

3.10 Conclusions

This chapter has presented a method for multimedia design based on a number of stages, each with a set of guidelines to be used by the designer. The method was based on the framework presented in chapter 2.

In order validate the guidelines, the following chapters 4 & 5 present studies which were motivated to test the guidelines for attentional design, media selection, and within media design. Since the guidelines were based upon a diverse set of literature, little of which directly reported studies on multimedia products, it was important to now test the guidelines for use in multimedia design. This was done by performing a set of empirical studies upon an existing multimedia product.

Chapter 4 reports a series of eye tracking studies which attempt to validate the importance of designing for attention, and to empirically justify a number of the guidelines presented.

Chapter 5 investigated attentional and within media design by testing comprehension. The guidelines were used to re-author the presentation to study if they would have real impact on what information was remembered from a multimedia presentation. Chapter 5 also provides a study of media selection, comparing the effectiveness of text & speech with full multimedia.

To test the usability and utility of the method, chapter 6 reports a study of how novice designers used the guidelines to author a multimedia presentation. Chapter 7 considers how the method can be encapsulated within an authoring and critiquing tool.

Chapter 4

Study : Eye tracking 'Etiology of cancer' presentations

4.1 Introduction

One of the key areas which the framework presented in chapter 2 suggested was the design of multimedia presentations for attention.

To investigate visual attention the study examined the order in which visual information is processed by the viewer using an eye tracking system. Of particular interest was the attentional processing of different visual presentation elements, such as text captions, labels and still and moving objects; and in how the serial nature of visual attention would cope with processing a complex and rapidly changing multimedia presentation. Few studies have made use of eye tracking equipment to look at animated materials, and none have attempted to track subjects viewing a complex multimedia presentation. However, visual attention is vital if visual information is to be successfully processed from the presentation.

4.2 Method

The pattern of fixations and attention upon the multimedia presentation were investigated using a Pupilometer eye tracking system. The objectives of the study were to attempt to match the shifting of fixations to the visual presentation elements shown, and to investigate any common patterns.

The studies were exploratory, in that they used existing materials which were not specifically designed to test any particular guideline from the method but investigated how attention reacted to the complex changing nature of an multimedia presentation. The choice of using a commercial multimedia presentation allowed the study to explore the effects of good and bad design in the real world, rather than making sequences which contained pre-defined problems. Parts of this study are reported in [Faraday & Sutcliffe 97].

Eight subjects were drawn from research students within City University. They were each pre-tested to ensure that they had no background knowledge of the subject area. They were informed of the aims of the study, and told that they would be shown each presentation once, and then asked question concerning it's content.

A Gulf & Western based pupilometer system was fixed upon the subjects eye, whilst they viewed the multimedia presentation. The eye tracker uses an infra-red light which is reflected from the pupil producing a bright spot which changes location as the eye moves. A video camera is then focused upon the bright spot (figure 4.2). The picture from the video camera is fed into a PC which produces an x,y location for the spot every 50 msec and filters out spurious data from subject blinks. Due to equipment problems, two subjects had to be discarded. The system had to be calibrated by getting the subject's to fixate on the four corners of the screen. Because any movement of the subject's head would cause the system to become uncalibrated a bite clamp was used to prevent head movement. They were held approximately 50cm from a 17 inch VGA colour monitor.

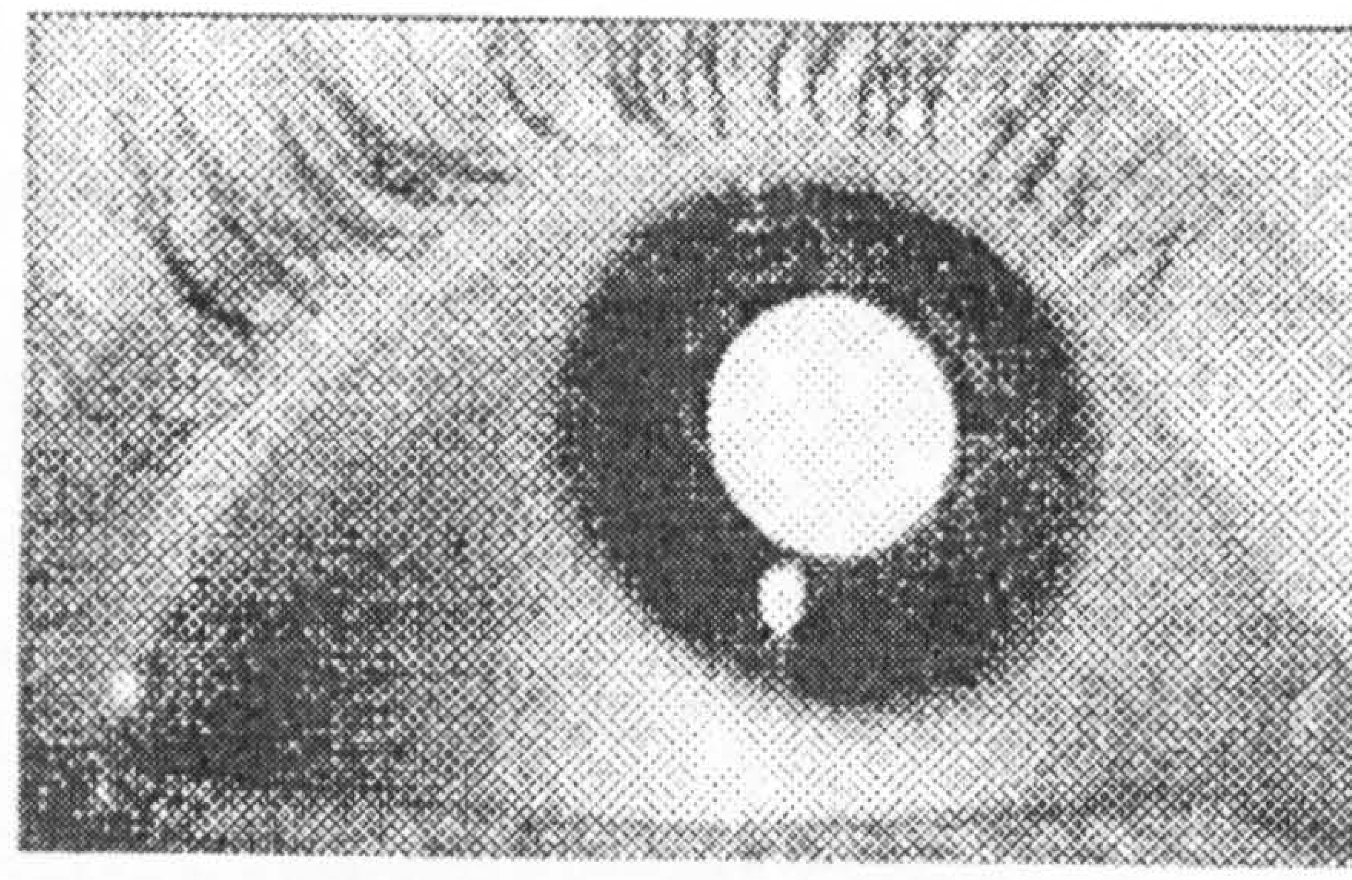


Figure 4.2 Bright spot on the pupil of the eye

4.2.1 Materials

The materials used for the study were taken from a commercially produced CD-ROM, 'The Etiology of Cancer' (figure 4.1). The product was designed to give GP or medical doctors an overview of how cancer develops in the human body.

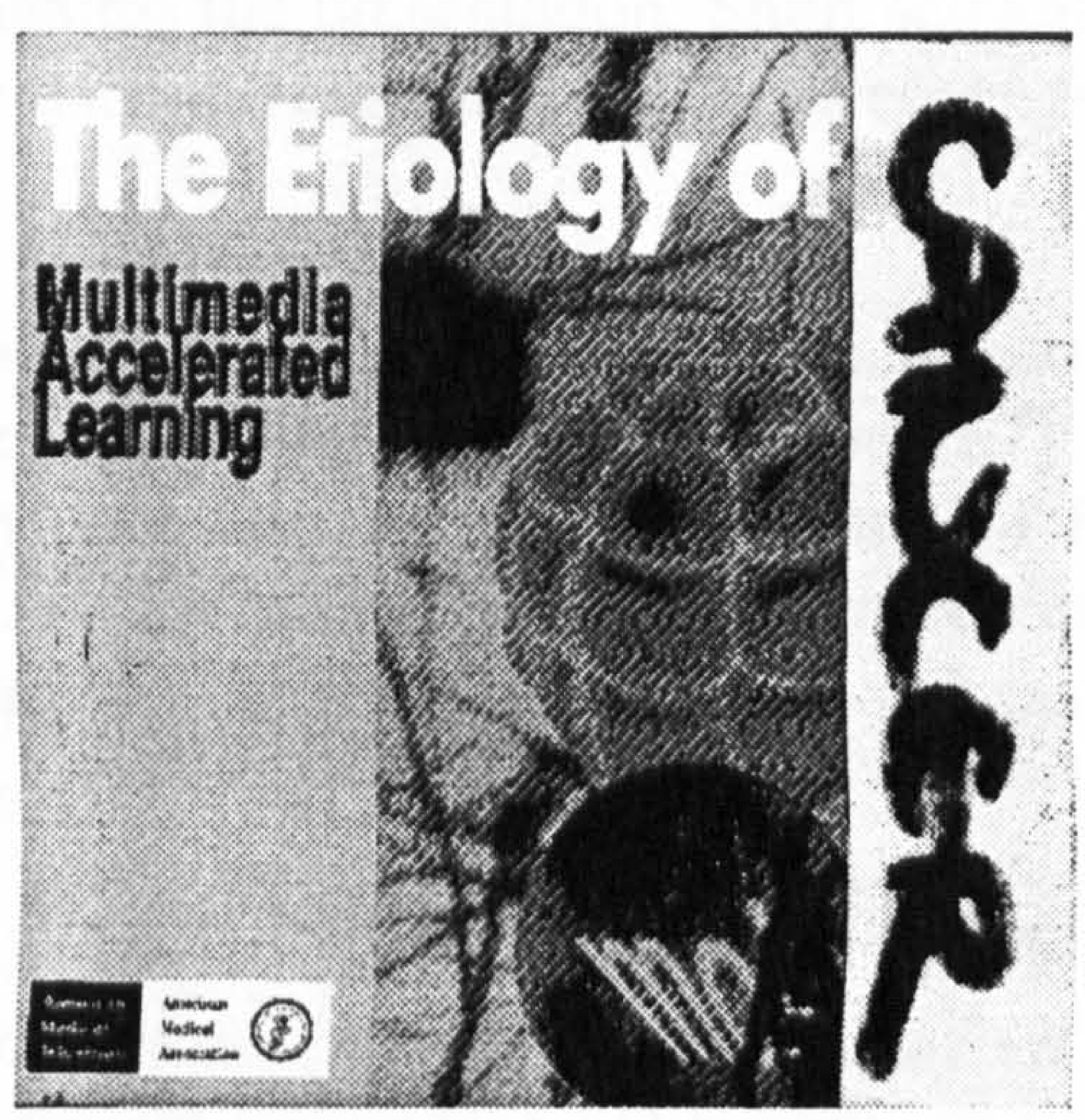


Figure 4.1 Etiology of Cancer product used in studies.

Three sequences were taken from the product : 'Metastasis', 'Ames Test' and 'Photoreactivation'. These are shown in sections 4.3, 4.4, and 4.5. Each sequence chosen contained a complex mixture of animation, speech and text. During the study, the subjects were shown the sequences prior to this material to ensure that they had sufficient context to understand it.

4.2.2 Data Analysis

The raw data of pupil location was time sliced into five second parts for each subject to ease analysis. A PC based tool was built to plot the data for each subject as a scan path. The data for the six subjects was synchronised over time to build up a composite trace for each part. This was placed upon a 32 by 24 grid so that it could be mapped to the 640x480 VGA screen resolution and compared with the multimedia presentation. The presentation screens in each sequence were then divided up into sets of objects.

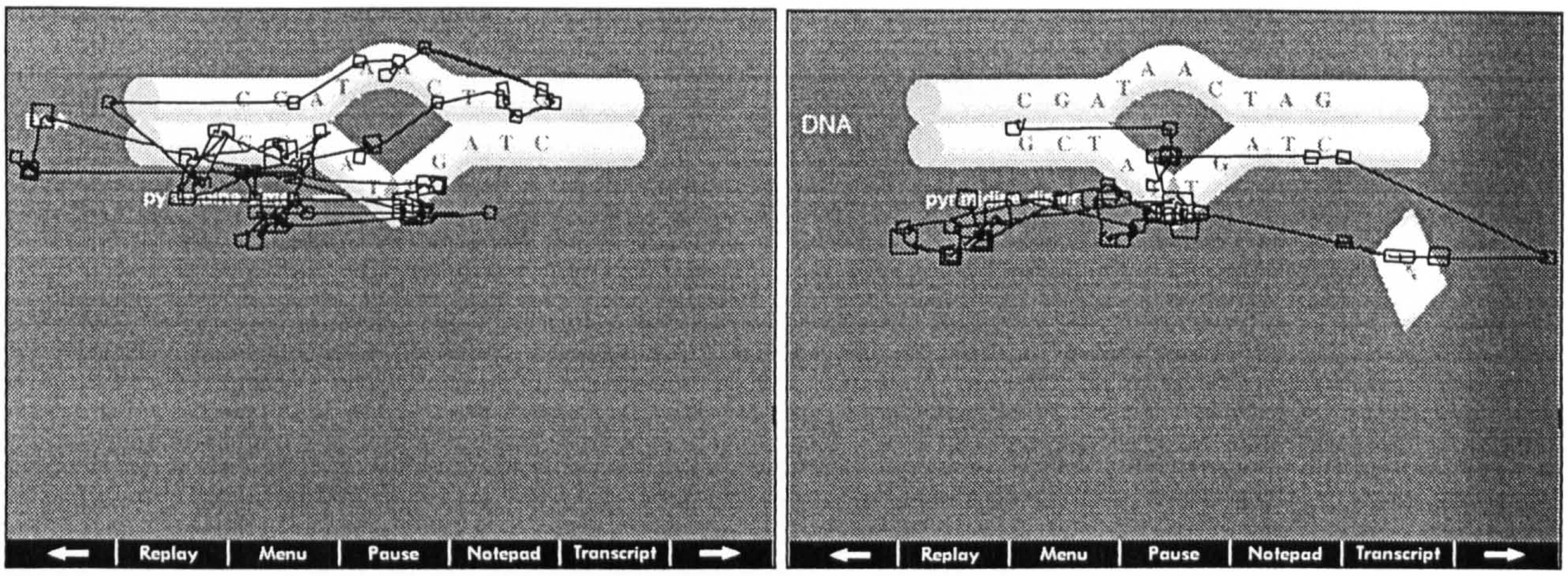


Figure 4.2.1 Single subject scan path

The fixation maps were compared by hand with the presentation sequence to produce a pattern for fixation clusters against the presentation. Short sequences of tracks were produced from fixed time points in the data and compared with the corresponding presentation at those points in time.

The decision of how many points from the eye track data should be scored as a fixation was based on known timings for visual identification. [Yarbus 67] has found that reading and object identification typically take over 200ms, so a cluster of four or more fixation points was scored as a fixation 'hit'. A 'hit' was ignored if its location could not be identified as one of the objects in the presentation. These errors were usually easy to spot, since they would show up as spikes in the data.

An example of the eye track data is shown in figure 4.2.1. Whilst it is known that pupil location and attention may not always coincide, in general a fixation, or pause, of the pupil at a location will indicate attentional processing. The parts of the presentation fixated were analysed to produce an aggregate of the shifts in fixation.

The scores for fixations were then combined for all of the subjects into a set of directed graphs. Figure 4.2.2 shows an example of the outcome of this analysis based on the sequence in figure 4.2.1.

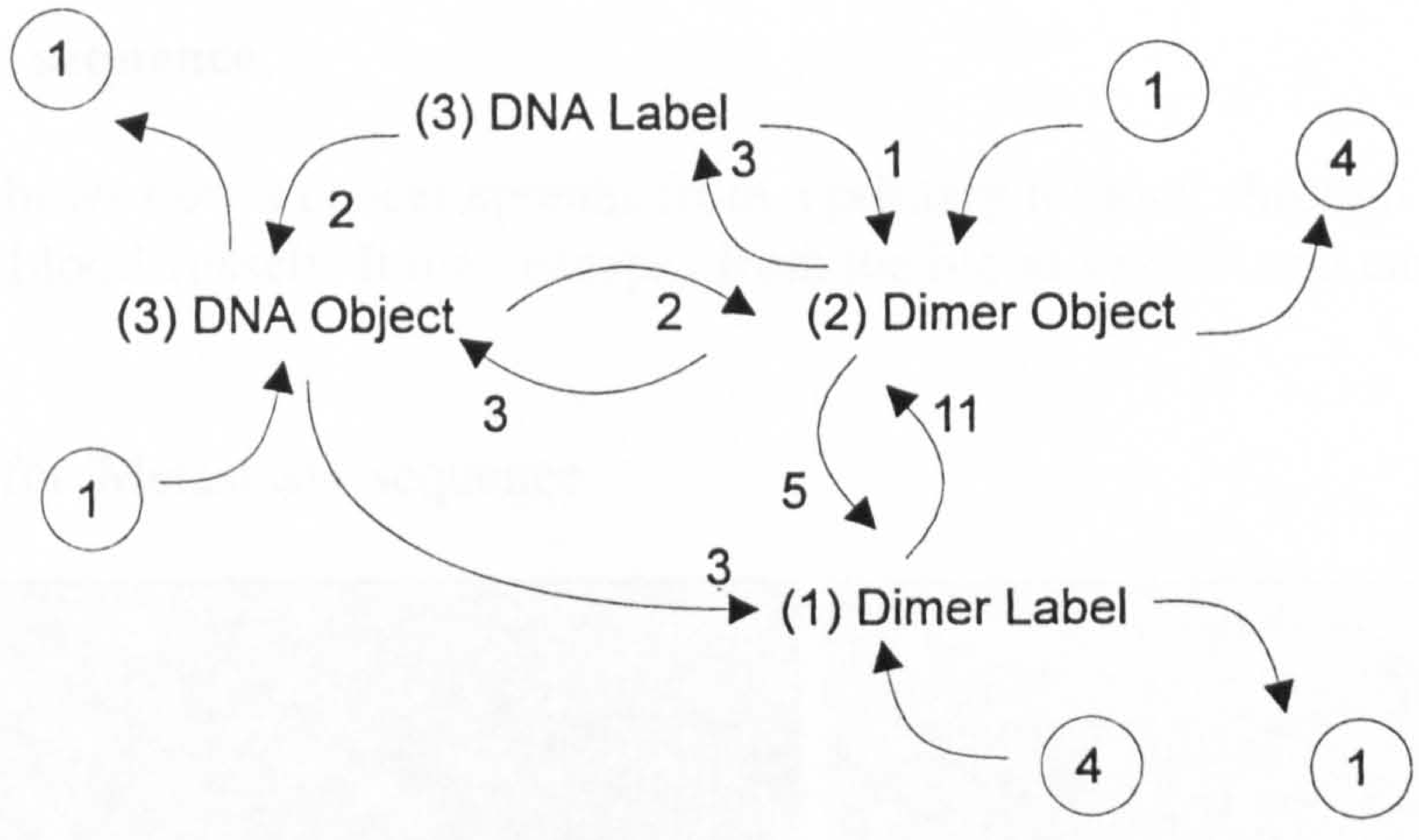


Figure 4.2.2 Example of notation used based on figure 4.2.1

First, the entry and exit points are shown on the graphs as circled numbers. This is where the subjects fixation either entered or left the presentation. For example, in figure 4.2 four subjects entered at the dimer label, and one exited the presentation there.

The arrows in the directed graphs show the direction of fixation. Thus the arrows show if the subjects entered or left the presenation, and to which part of the presenation they then went.

The number at the end of each arrow shows how many subjects shifted from one part to another. If subjects shifted back and forth between parts, the number will be larger. An example of this can be seen for the dimer label in figure 4.2.2 in which eleven fixations shift to the dimer object, and five fixations shift back to the dimer label again.

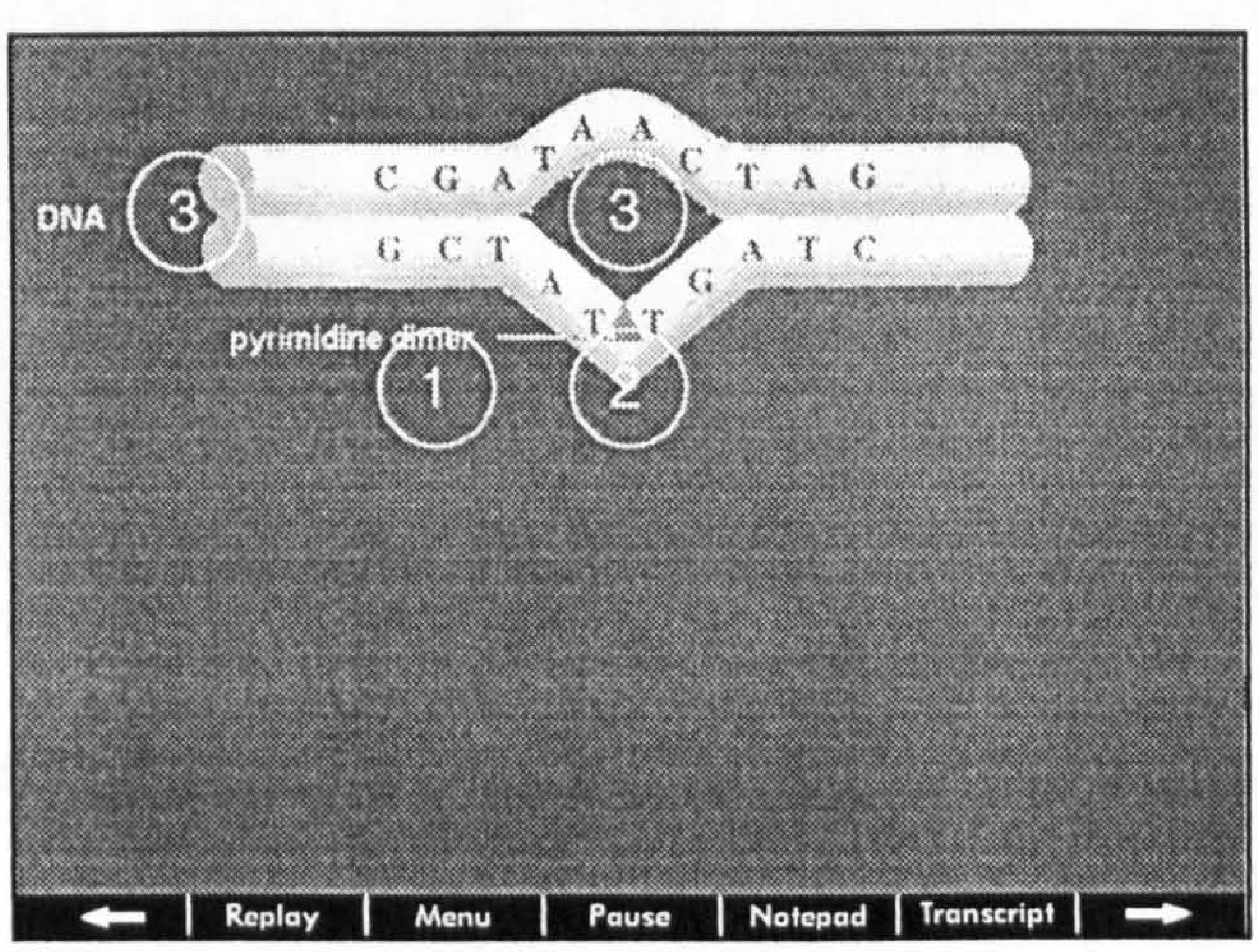


Figure 4.2.3 Example of screen shot annotated with fixations

To help understand how the elements in the graph are related to what happened on screen, the circled numbers on the screen shots below indicate the gross order of fixations indicated by the graphs. An example is shown in figure 4.2.3. The gross order is from the dimer label, to the dimer, and then to the DNA and it's label.

4.3 'Metastasis' sequence

This sequence shows how a cancer spreads from a primary tumour, through the surrounding tissues, into the blood vessels. It then escapes from the blood vessels and causes secondary tumours.

4.3.1 Materials for 'Metastasis' sequence

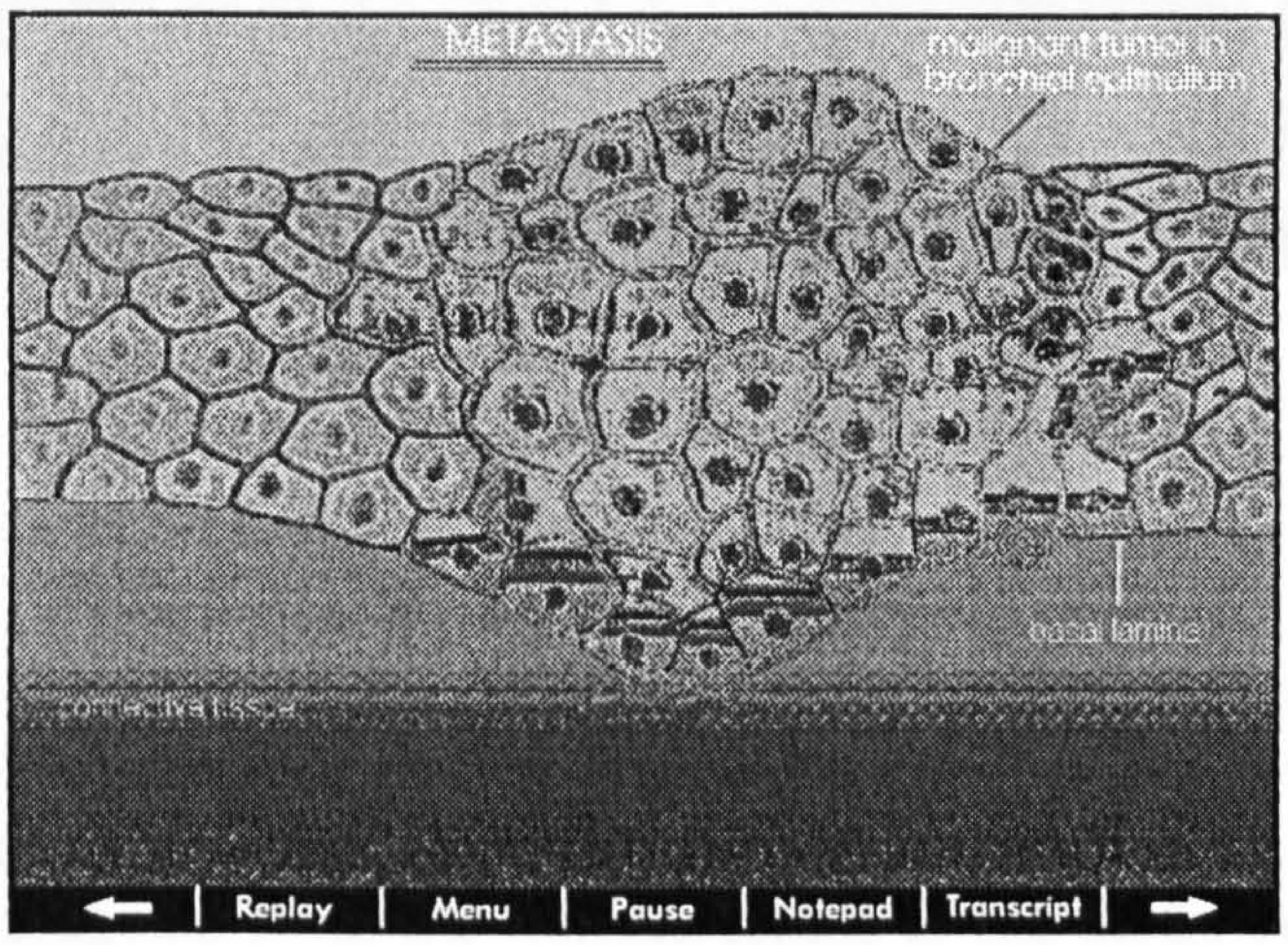


Figure 4.3.1

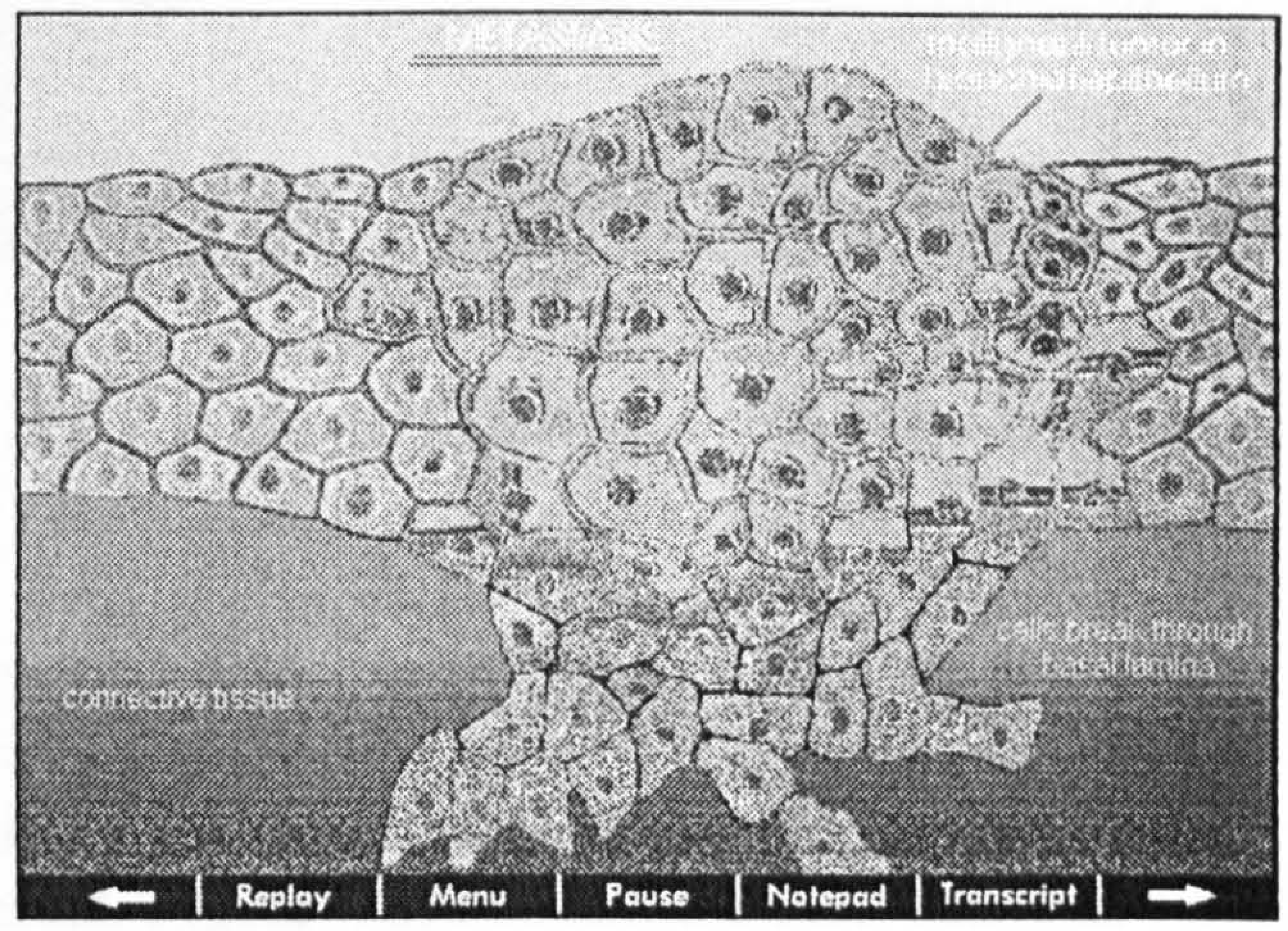
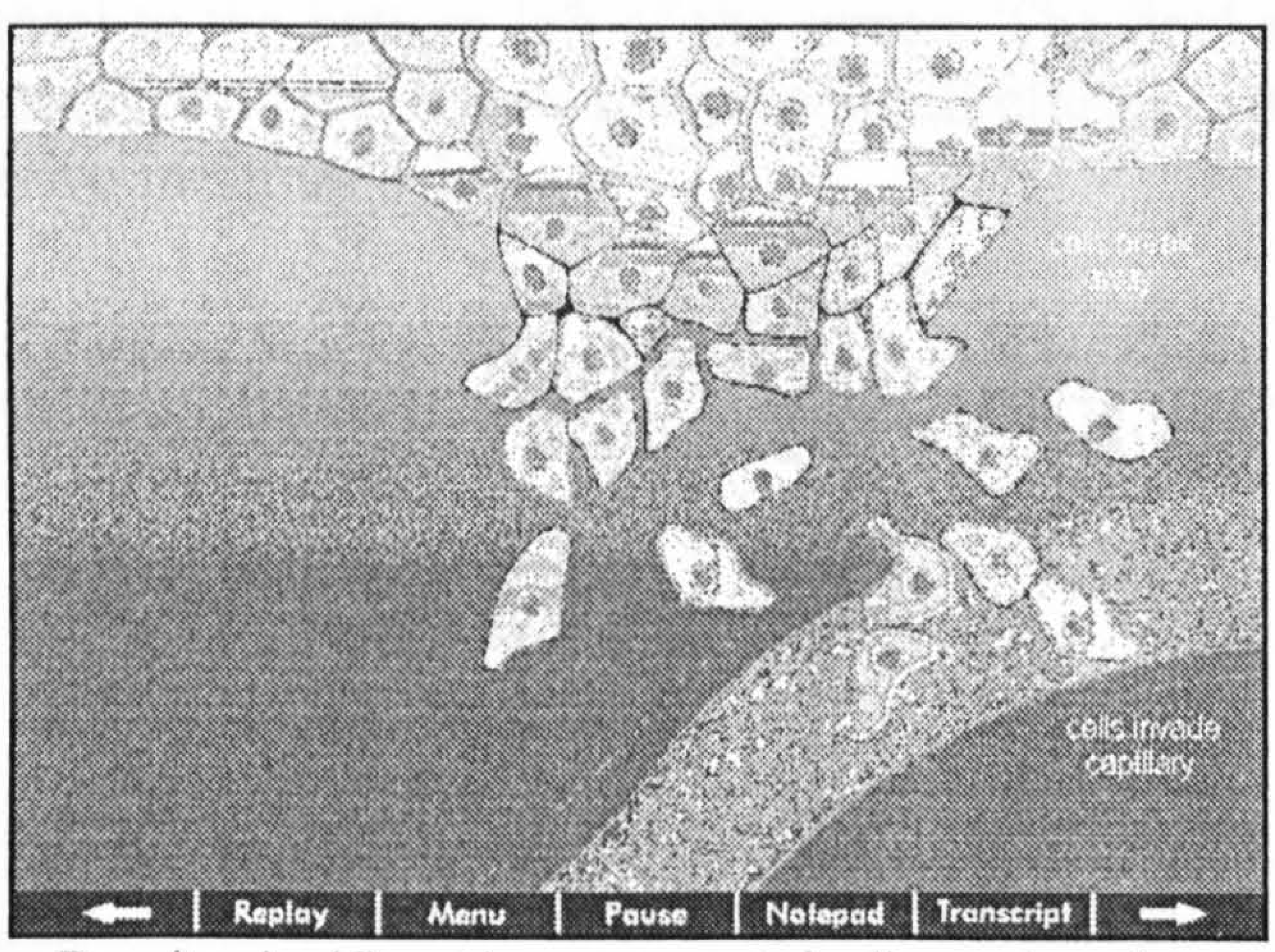
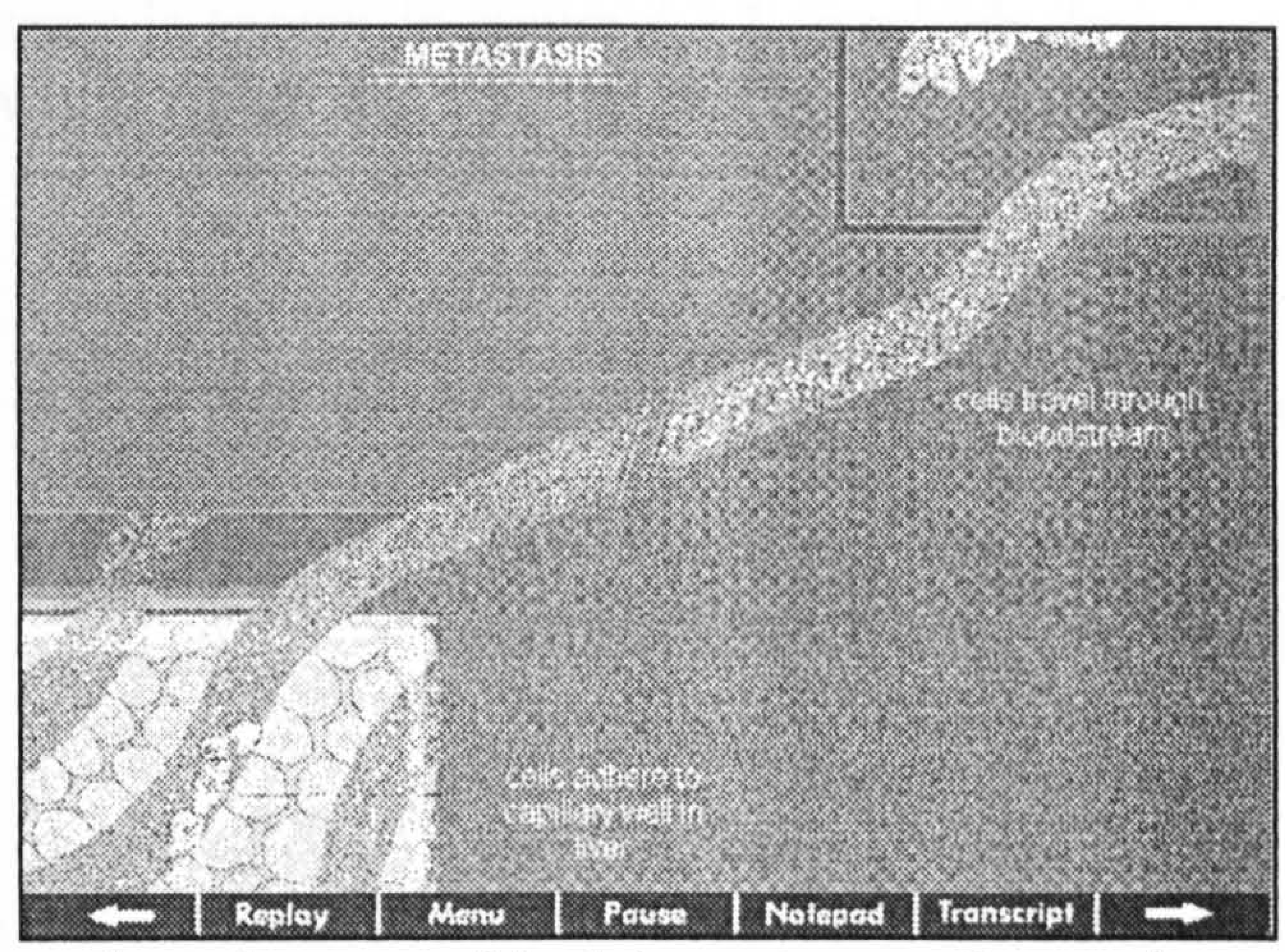


Figure 4.3.2



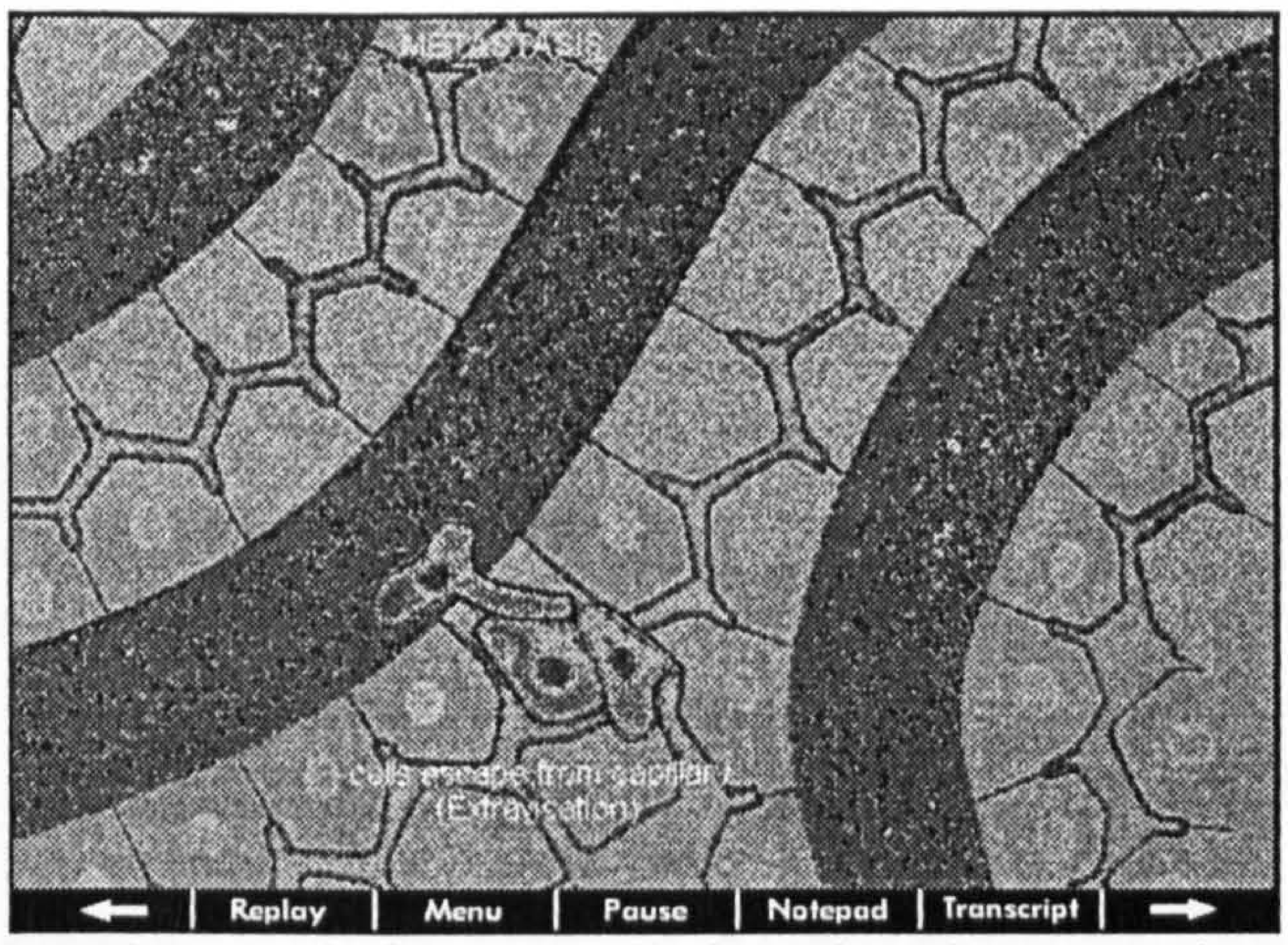
Despite significant improvements in diagnosis and treatment ...

Figure 4.3.3



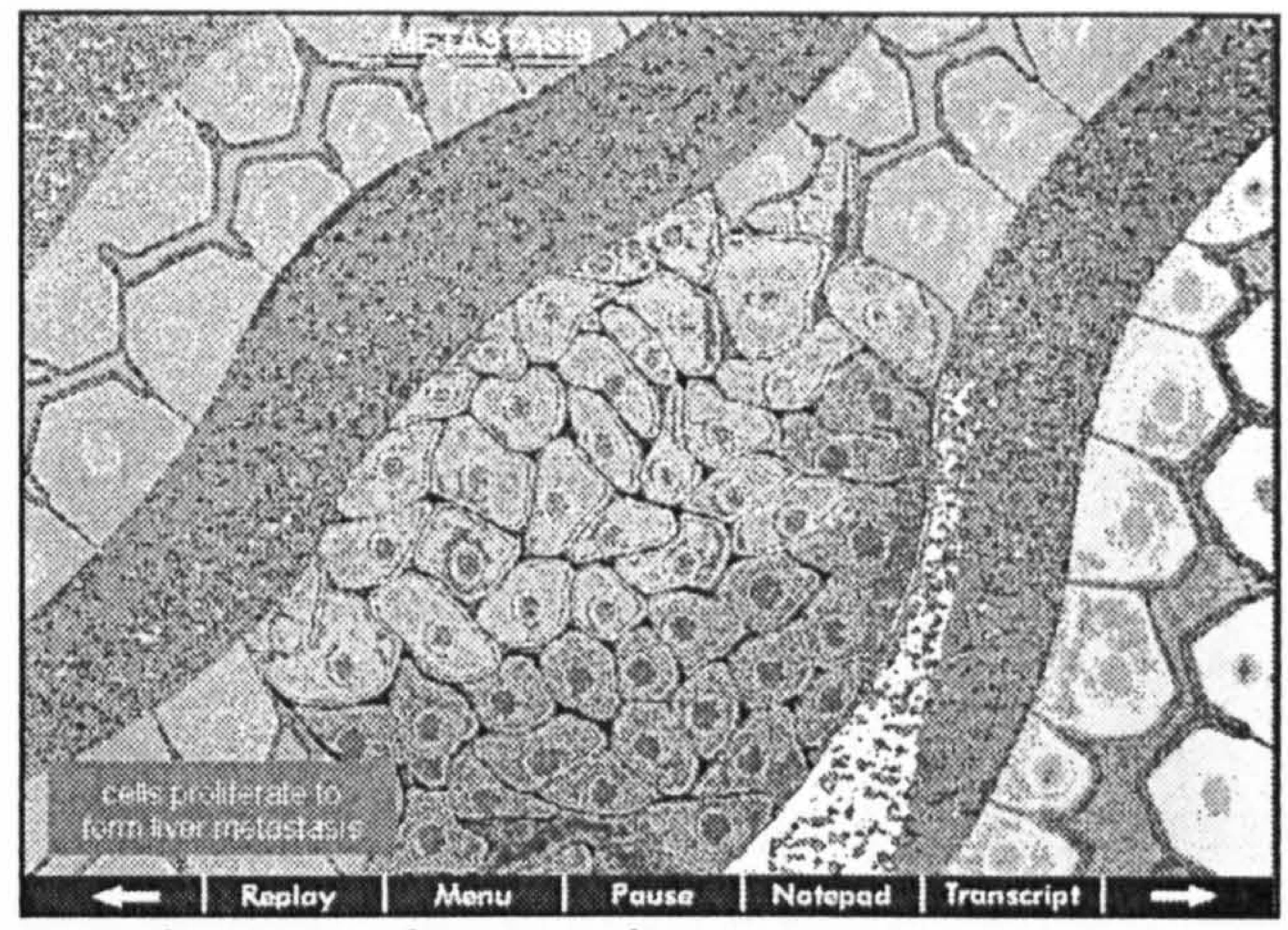
... the major cause of death in most cases of cancer is metastasis ...

Figure 4.3.4



... the spread of tumour cells from the primary neoplasm to distant organs ...

Figure 4.3.5



... where secondary growth occurs.

Figure 4.3.6

In figure 4.3.1 the malignant tumour object in the centre of the screen is shown moving down toward the bottom edge of the screen. The scene ends as shown in figure 4.3.2. A new scene begins in figure 4.3.3 with the tumour breaking up into a series of tumour cells which move in a diagonal from the tumour across to the blood vessel. Figure 4.3.4 shows the next scene in which tumour cells move from the top right of the screen in a diagonal following the blood vessel to the bottom left of the screen. The final scene is shown in figure 4.3.5 in which the tumour cells proliferate from the small cluster shown in figure 4.3.5 to the secondary tumour shown in figure 4.3.6.

4.3.2 Results

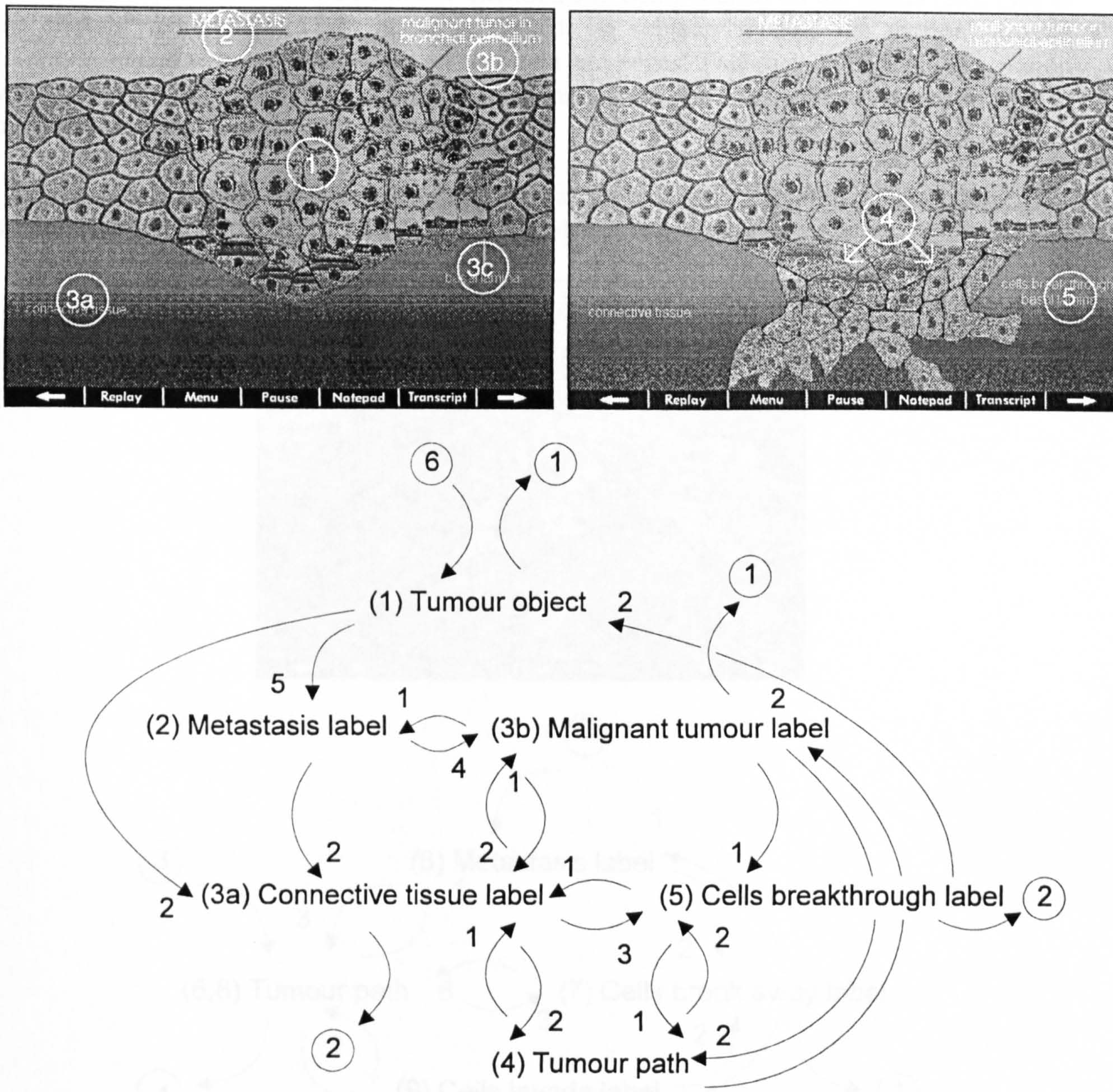


Figure 4.3.8

(1) All subjects initially produced orientation fixations, beginning by transversing over the large central tumour object presentation.

(2) Five of the subjects then shifted to fixate upon the region of the 'Metastasis' title label

(3a, 3b, 3c) Four subjects shifted to the 'malignant tumour in bronchial epithelium' label, one subject to 'connective tissues' label. No particular order was found across the subjects. The 'basal lamina' label (3c) was not fixated upon. This may be because it was removed shortly after the start of the presentation and replaced with label (5) during the animation (4).

(4) Animation of the malignant tumour cell spreading downwards. The motion caused a general shift of fixation to the centre of the presentation; but the shift did not happen at the start of the motion. This could be due to attention still being locked on the labels (3a), (3b) at the start of the motion. Once attention shifted, no subjects tracked the motion, instead fixations tended to be given upon its general area and edges. This may be because the motion was shown for only a short period, and having missed the start, the subjects were not able to track it.

(5) Revealed caption 'Cells break through basal Lamina' concurrent with animation. All subjects tended to fixate upon this, as it was revealed.

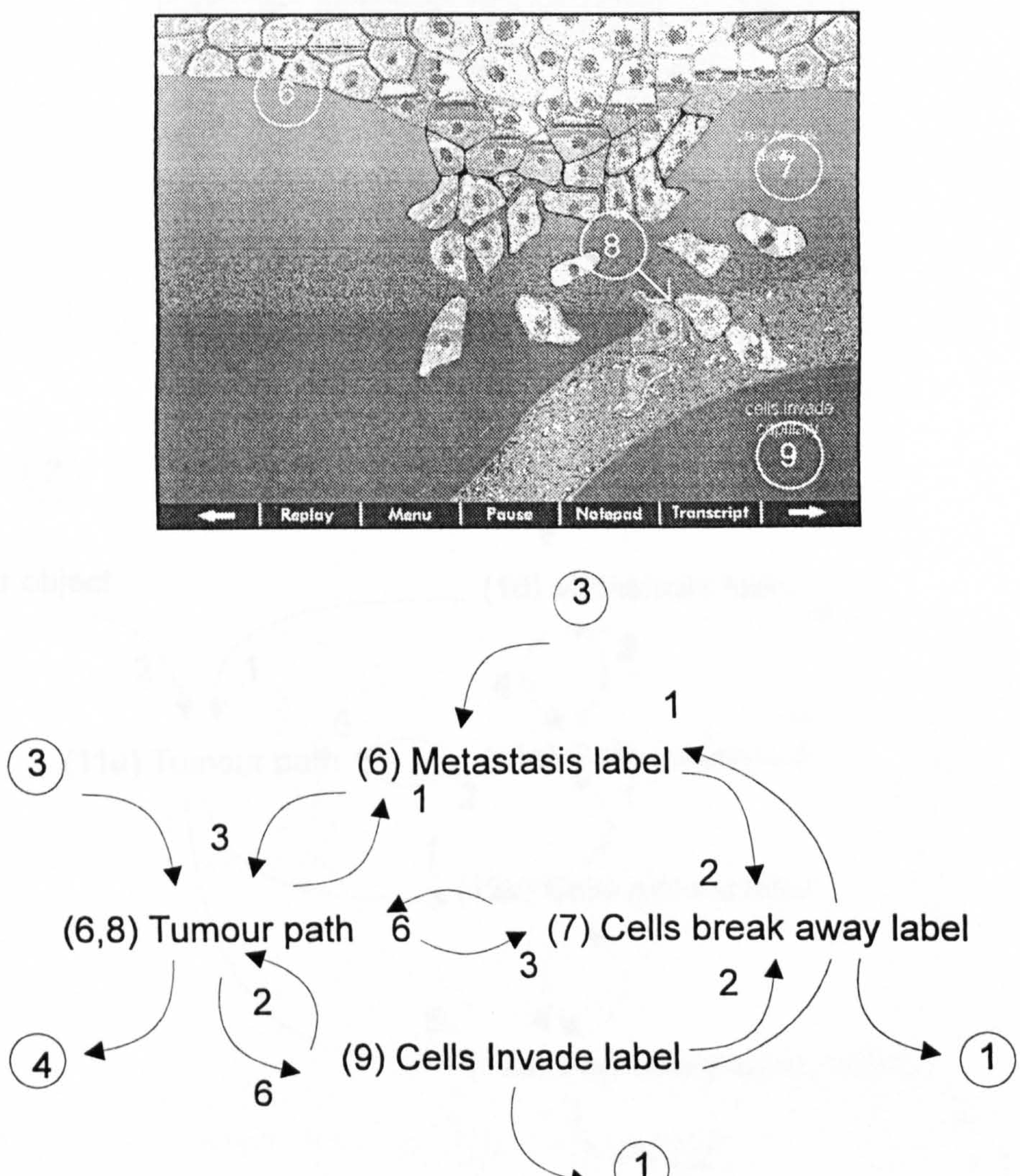


Figure 4.3.9

(6) The scene was changed. Three subjects shifted to the 'Metastasis' label, and three directly to the tumour path. The tumour path showed the tumour object away from its rest position initially.

(7) All subjects shifted to 'Cells break away' label when it was revealed.

(8) Animation of the tumour cells moving in a diagonal from top-middle to bottom-right. All subjects showed an alignment on the tumour cells, with subsequent fixations tracking their path of motion.

(9) All subjects shifted to the revealed caption 'Cells invade capillary' at the end of the path of (8). Two subjects then shifted fixations back and forth between the tumour cell objects and two subjects shifted to 'Cells break away' label.

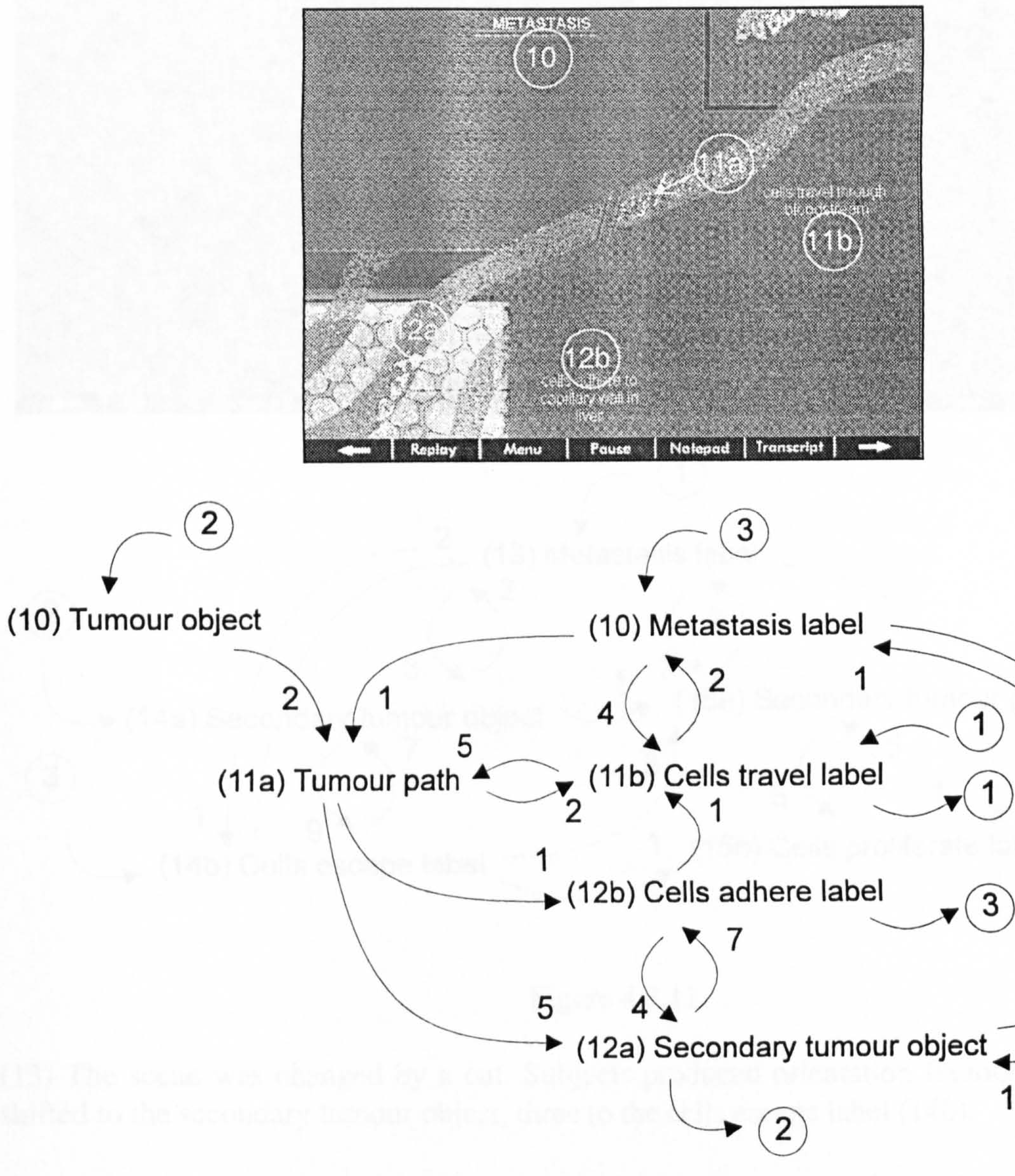


Figure 4.3.10

(10) The scene was changed. Subjects produced orientation fixations and then three shifted to the title 'Metastasis', and two to the tumour object itself.

(11a) Animation of the tumour cells moving in a diagonal from top-right to bottom-left. No subjects fixated upon the start of the motion, only picking it up at about 1/4 through its path. Instead fixations tended to be either still on (10), or on (11b), before picking up the path.

(11b) Static label 'Cells travel through blood stream'. Three subjects shifted from (10) to (11b), and then to the animation (11a).

(12a, 12b) Five subjects shifted to secondary tumour object at the end of the path (11a) and then to revealed label 'Cells adhere to capillary wall'. All subjects shifted fixations from the tumour cells and to the label.

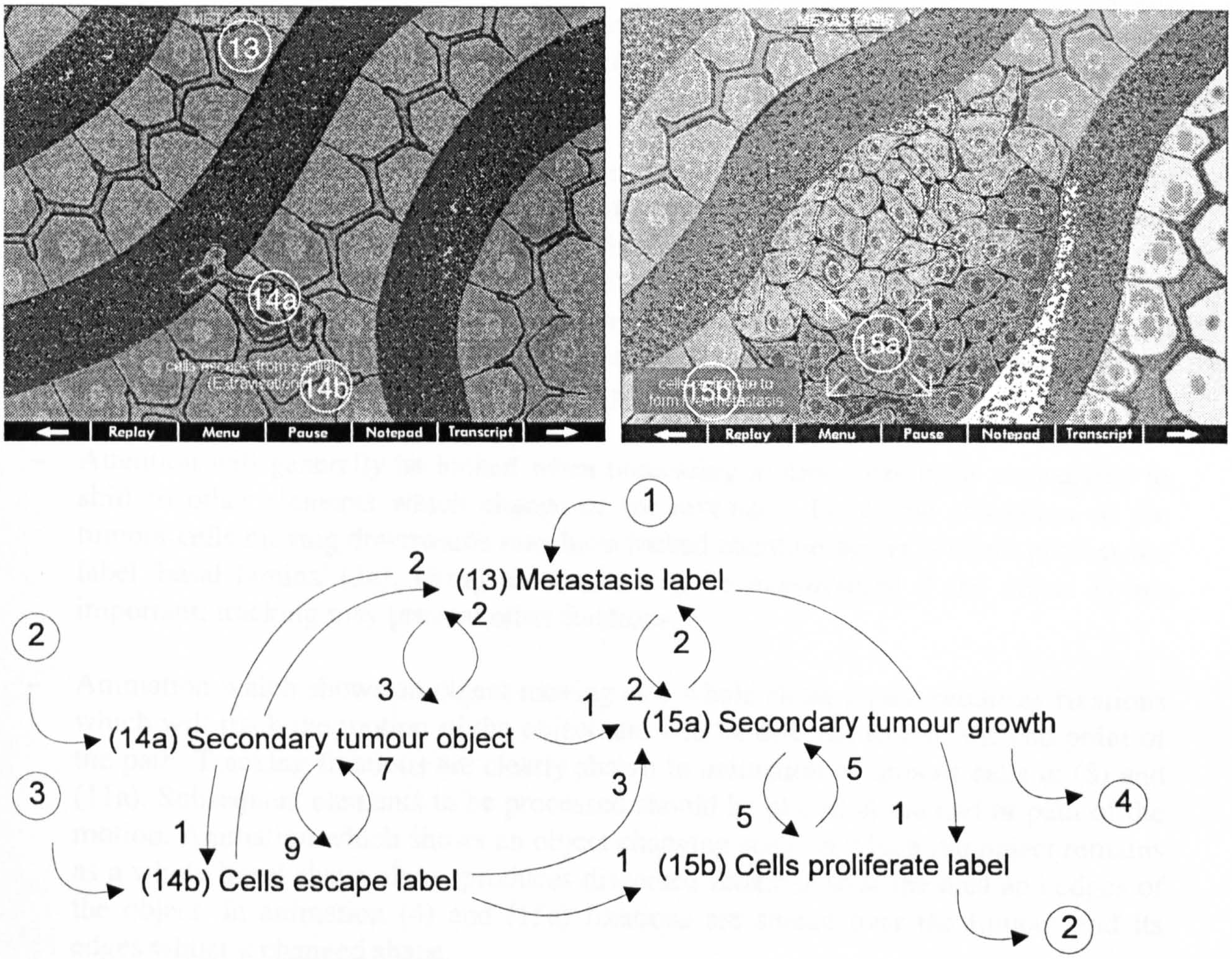


Figure 4.3.11

(13) The scene was changed by a cut. Subjects produced orientation fixations and then two shifted to the secondary tumour object, three to the cells escape label (14b).

(14a, 14b) Five subjects shifted fixations back and forth between the static secondary tumour object and static label 'Cells escape from capillary : Extravisation'.

(15a) Animation of secondary tumour cell growth. Five subjects fixated upon its general area and edges. The animation was internal, changing the object shape, and again did not produce tracking fixations.

(15b) Revealed label 'Cells proliferate to form liver metastasis'; five subjects shifted fixations back and forth between the secondary tumour growth object and label.

4.3.3 Discussion

The study provided some evidence for the effect of design techniques upon fixation and attention within presentations. Several conclusions may be drawn :

- Animation has a strong effect on attention. The onset of an animation will generally produce a shift of attention to the object in motion if attention is not already engaged elsewhere in the presentation. Evidence for this can be found in animations (8) and (15a).

  However, in animations (4) and (11a) the start of the animations were generally not shifted to immediately. In animation (4) subjects did not seem to shift to the downward spread of the tumour cells until late in their motion; whilst in (11a) the path of the tumour cells was not tracked by any subjects until the animation was around 1/4 complete. Instead attention may have been locked on the title 'Metastasis' (2), (10) and labels (3a, 3b, 3c) and (11b) in the two frames. If the start of the motion is important, then ensure that attention will be free, not processing labels or captions.

- Attention will generally be locked when processing motion, making it unavailable to shift to other elements which change or are revealed. In (4) the animation of the tumour cells moving downwards may have locked attention before it could process the label 'basal lamina' (3c). Using animation may be unwarranted if the object is not important; tracking may prevent other fixations.

- Animation which shows an object moving as a whole along a path produces fixations which will track the motion of the object and will be directed toward the end point of the path. Tracking fixations are clearly shown in animation of tumour cells in (8) and (11a). Subsequent elements to be processed should be placed at the end of path of the motion. Animation which shows an object changing state, in which the object remains as a whole but it shape alters, produces dispersed fixations over the area and edges of the object. In animation (4) and (15a) fixations are spread over the tumour and its edges whilst it changed shape.

- Changing scenes causes attention to re-orientate. Reorientation fixations were found after scene changes in (2), (6), (10) and (13). Each time attention seemed to process the scene anew, e.g. refixating on the previously attended title 'Metastasis' after each scene change. Using a cut in film tends to enhance attention to the material following the cut. Avoid revealing labels or object motion to allow re-orientation to take place after changing to a new scene.

- By gradually revealing labels, symbols and objects, fixations can be shifted e.g. between the tumour and labels in (5), (9). Reveal may be used to order attention from one part of the topic to the next. Revealing several presentation elements simultaneously may confuse the ordering; reveal important elements singularly.

- Static object elements are not attended to uniformly. Those which are labelled, which are bright in colour, or appear to be different from their surroundings will be attended to in preference to other objects. Use strong colours or labels to draw attention. Avoid using multiple labels, such as in (3a, 3b, 3c) which may lead to unpredictable shifting of attention.

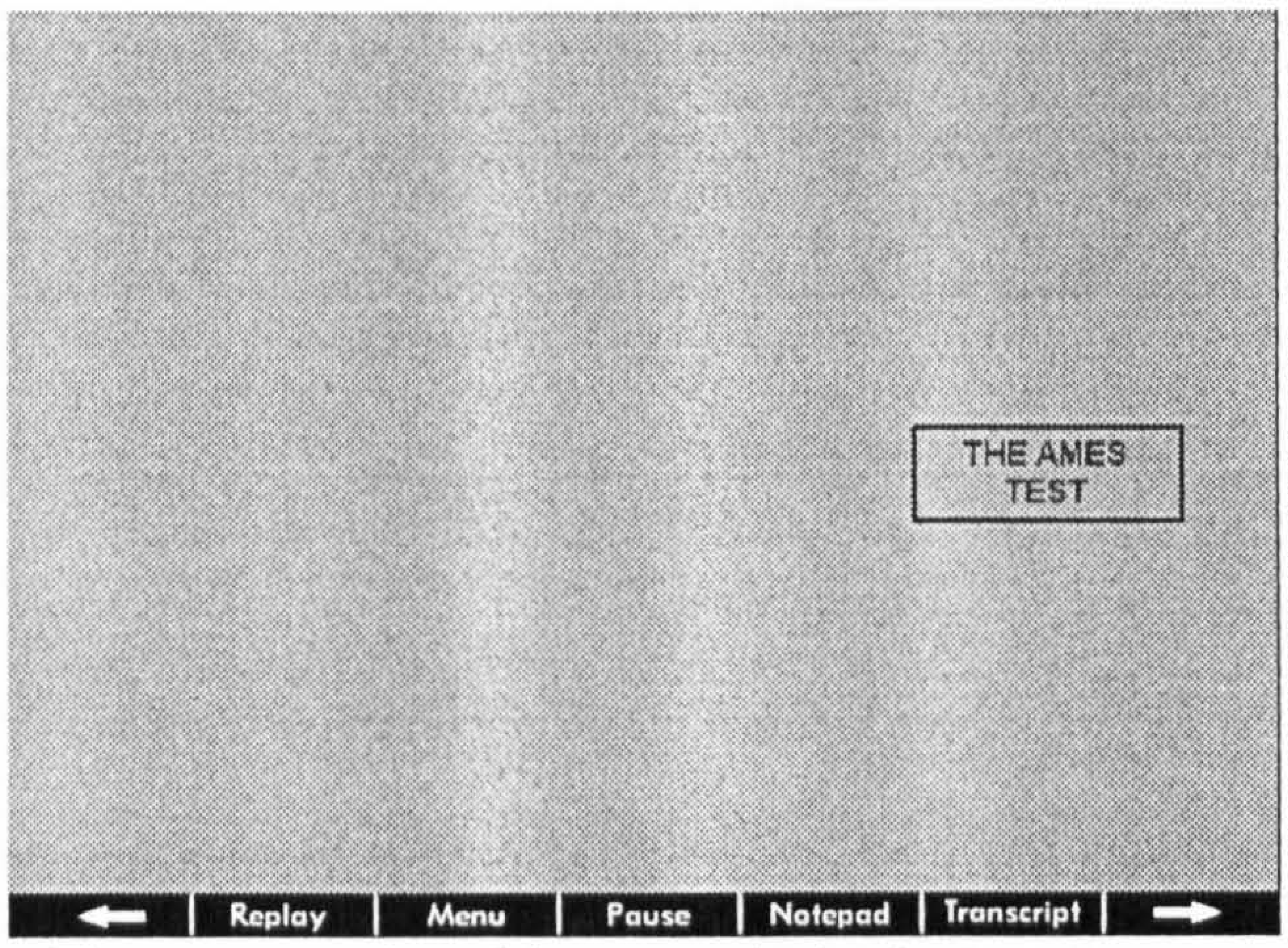
- Reading captions and labels will lock attention, preventing other elements being processed. Reading also requires longer fixation time than for identifying objects. Attention may have been locked on the title 'Metastasis' (10) and caption (11b), preventing processing of the start of animation (11a). Allow time for text elements to be processed.

- Labelling an object will produce fixations shifts between the object and label e.g. (13a,b) and (15a,b). The findings of [6] are similar suggesting that switches of attention are used to integrate information from the label with the object. Ensuring that an object and its label are both available together may aid this process.

4.4 'Ames test' sequence

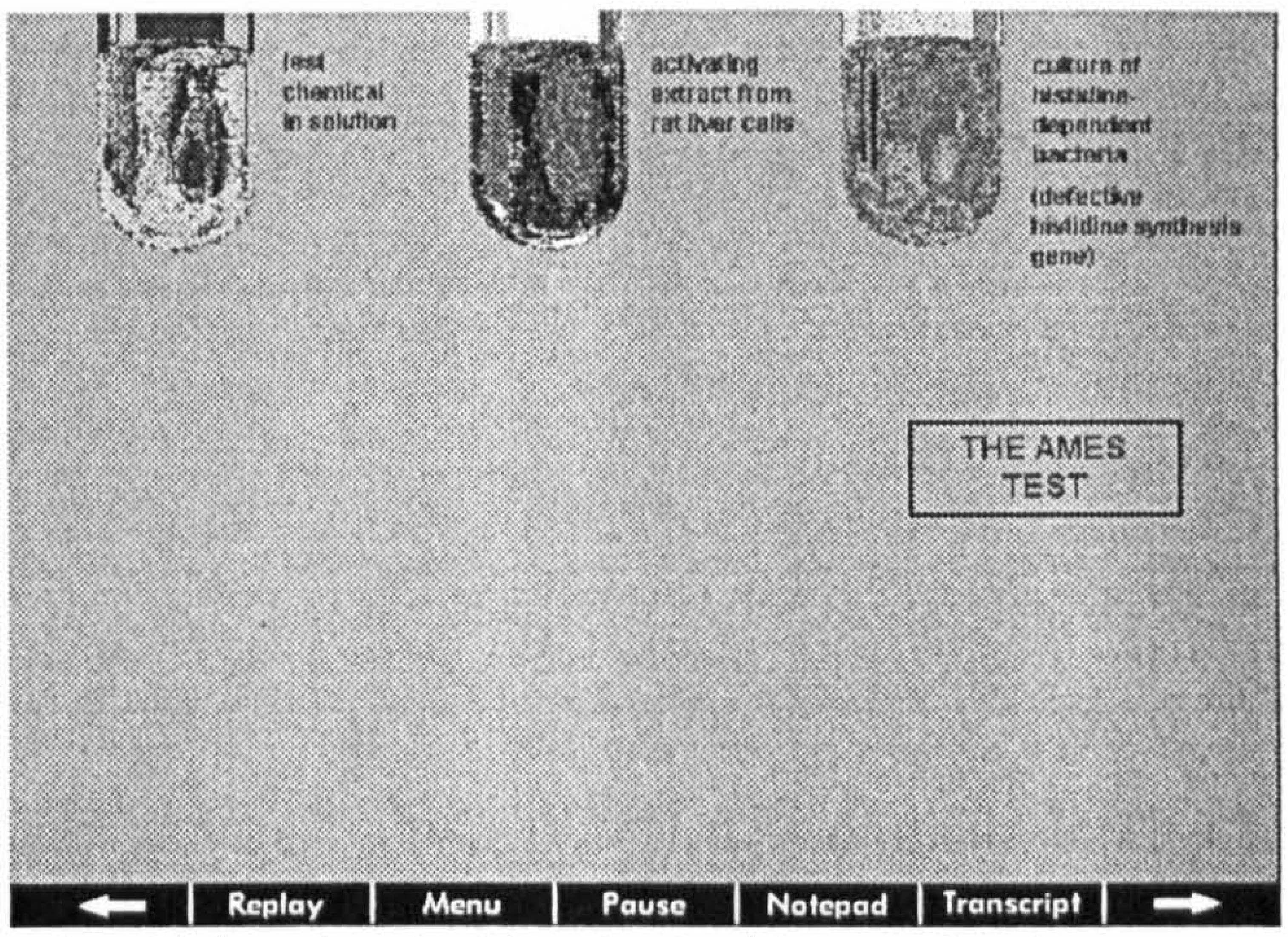
The 'Ames test' sequence illustrates how chemicals can be checked to see if they are a potential carcinogen. It makes use of the fact that if a chemical causes a mutation in cells, it is likely to be carcinogenic. The chemical to be tested is mixed with a specially designed bacteria which has a defect preventing it from growing. The mixture is then incubated, during which time the bacteria can only grow if the chemical causes a mutation. The amount of growth can then be measured. The more growth found, the greater then mutagenic and therefore carcinogenic potential of the chemical.

4.4.1 Materials for 'Ames text' sequence



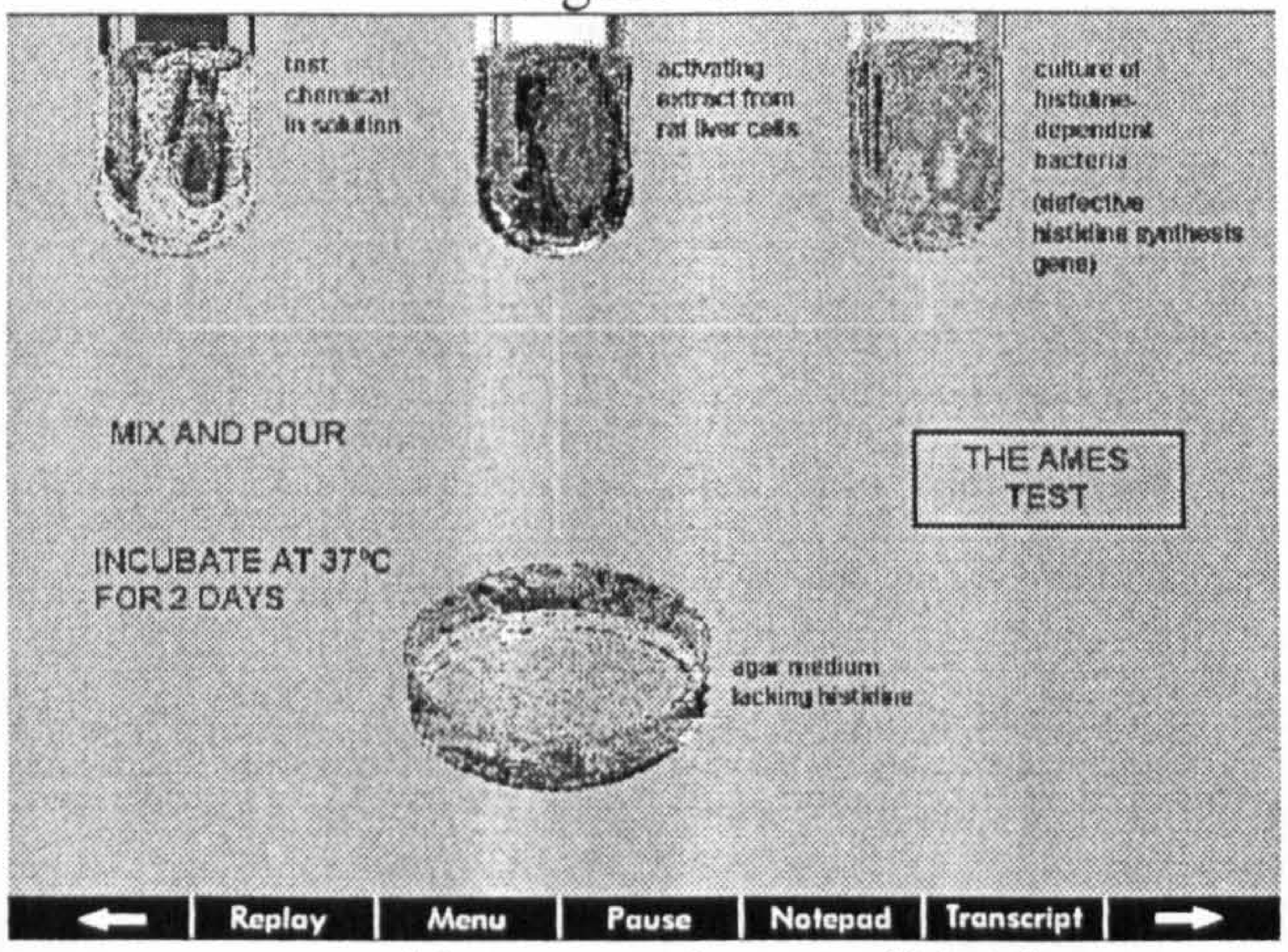
The most popular of these tests is the Ames test ..

Figure 4.4.1



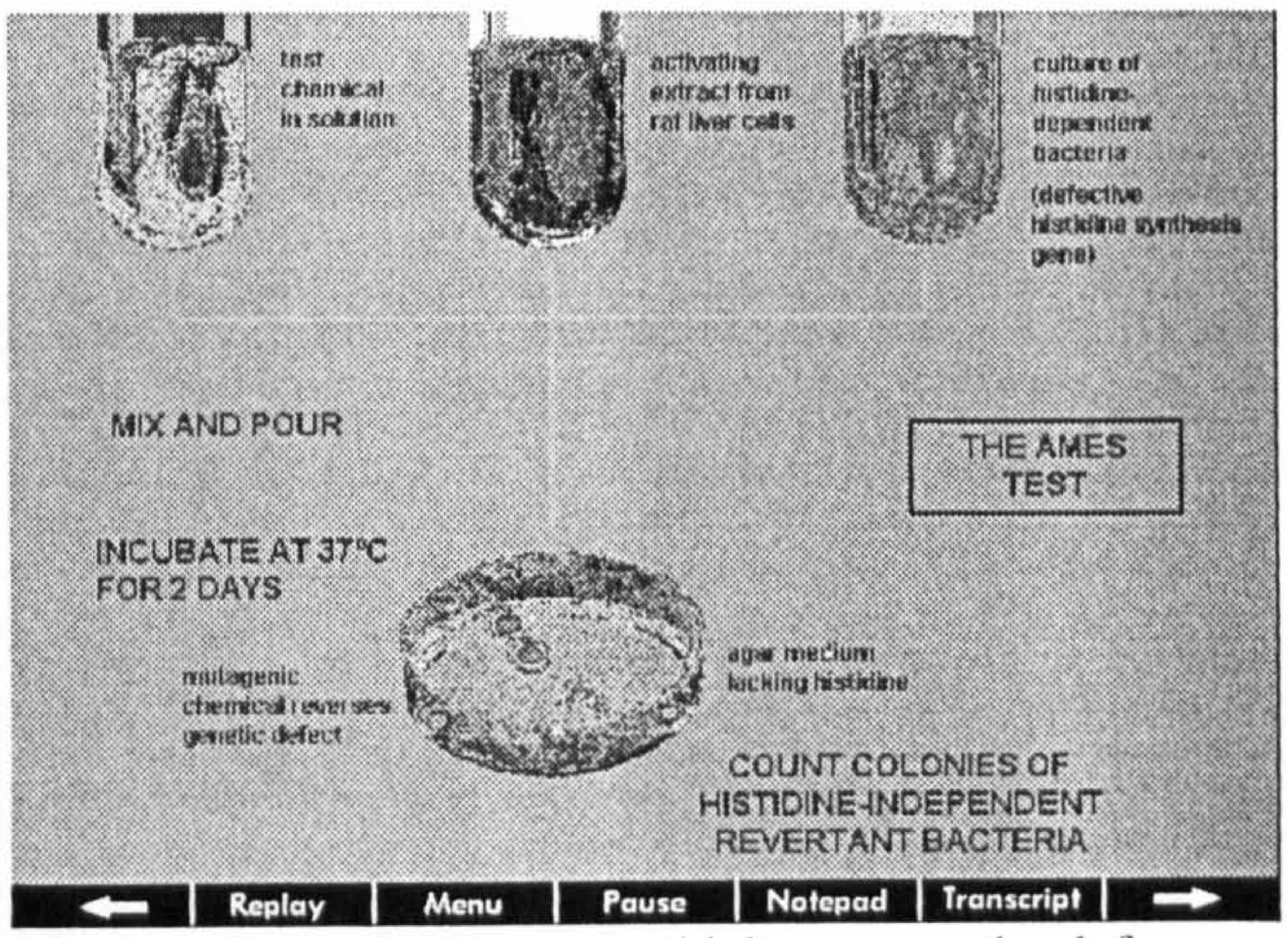
in which a chemical is mixed with an activating extract prepared from rat liver cells and added to a culture of specially designed bacteria which...

Figure 4.4.2



..cannot grow without a supplement of histidine, because they have a defect in a gene required for histidine synthesis. A mutagenic chemical can cause a..

Figure 4.4.3



further change in this gene which reverses the defect, creating reverant bacteria which will grow without a histidine supplement and which can be counted as a measure of the mutagenic and therefore carcinogenic potential of the chemical under test.

Figure 4.4.4

Figure 4.4.1 shows the start of the sequence, which introduces the ames test. Figure 4.4.2 shows the components of the text, which are revealed one at a time from left to right Figure 4.4.3 begins with the arrow symbol being revealed along with the label 'mix and pour' and the dish containing the test mixture. This is then followed by the label 'agar medium lacking histidine' being revealed. After this, the caption 'incubate at 37c for 2 days' is revealed. Figure 4.4.4 starts with an animation of the cells in the dish changing colour and the reveal of the label 'mutagenic chemical reverses genetic defect'. Finally the caption 'count colonies of histidine independent revertant bacteria' is shown.

4.4.2 Results

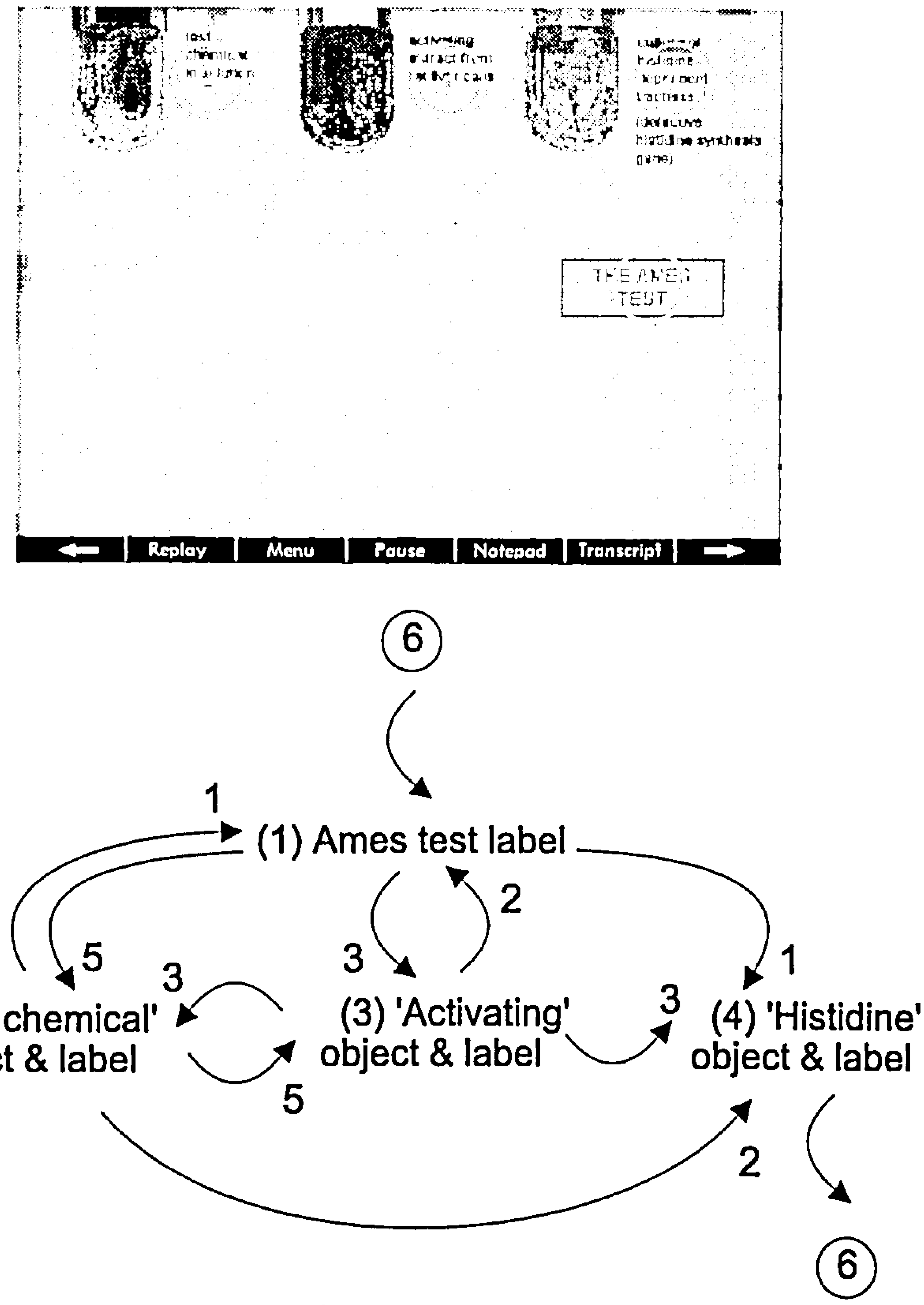
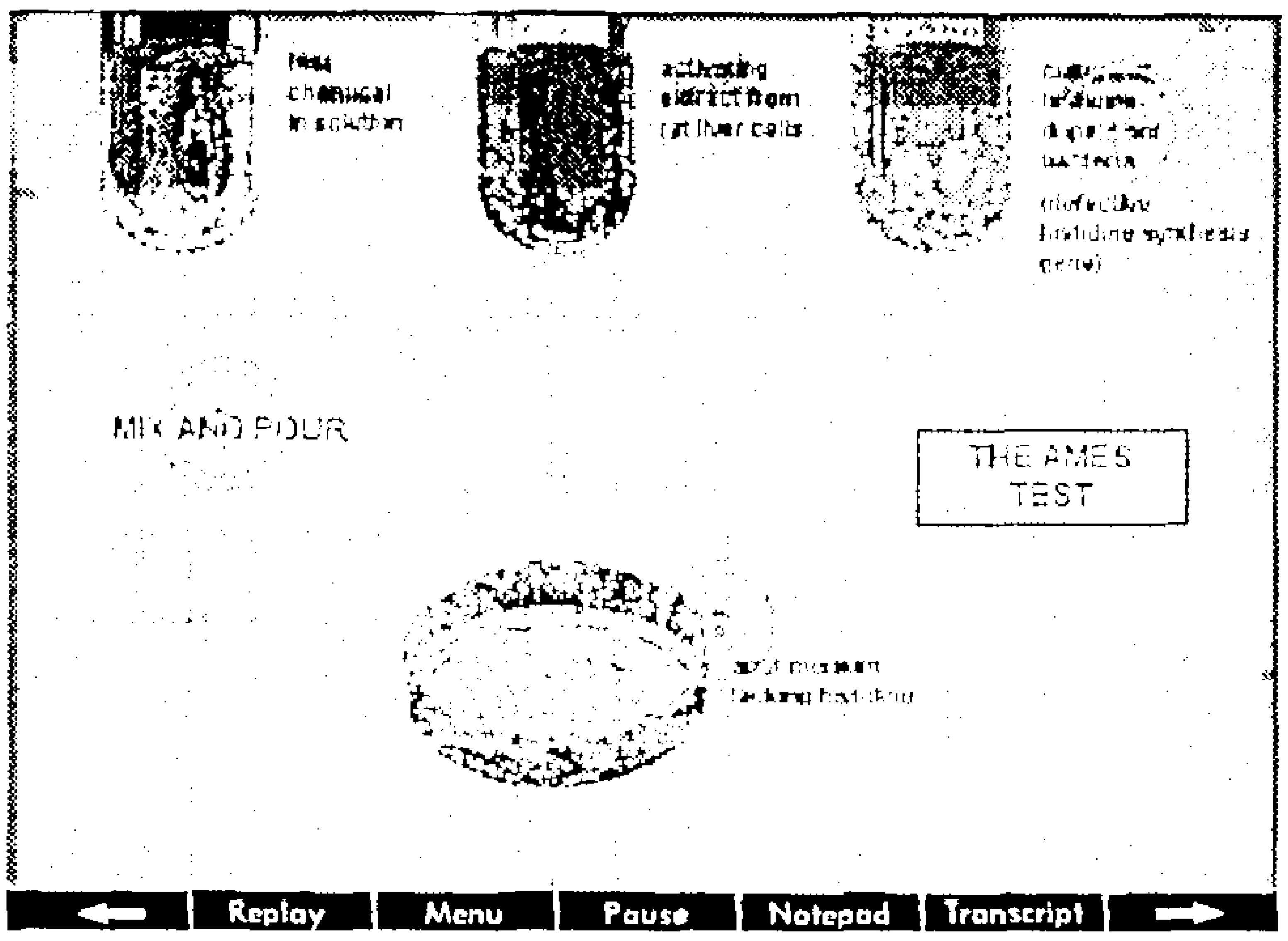


Figure 4.4.5

(1) All six subjects fixated on the ames test label. This was label was shown on its own.

(2), (3), (4) Revealing the test tube objects and labels in series shifted fixations from (2) to (3) to (4). Some refixations were also found from (3), (2), which may have been caused by the speech referring to the objects as being 'mixed' together. All subjects fixations ended this sequence on the 'histidine' label (4)



6

(4) 'Histidine'
object & label

6

(5) 'Mix and pour' label

1

8

(6) Culture object
& 'Agar' label

6

Figure 4.4.6

(5) All subjects shifted from 'histidine' label (4) to the 'mix and pour' label (5), which was revealed.

(6) The 'mix and pour' label was accompanied with an arrow symbol, pointing towards the 'culture object'. This may have driven fixations down the presentation to (6).

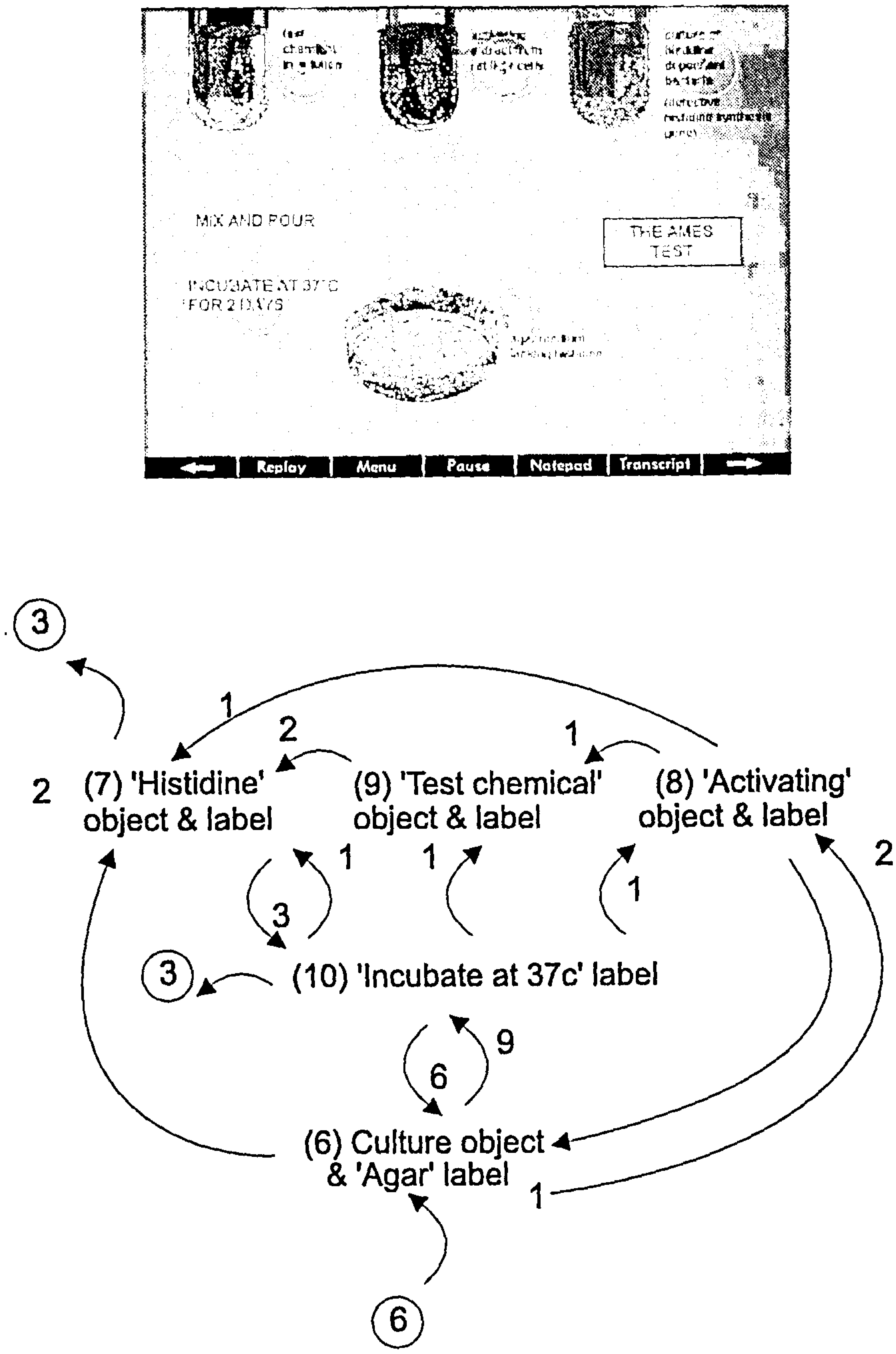


Figure 4.4.6

(6) All Subjects shifted to the revealed culture object and 'agar' label. Three subjects shifted back and forth from the object to the label

(7), (8), (9). Four subjects then refixated on the 'histidine' object / label (7), 'activating' object / label (8) or 'test chemical' object / label (9) ; this may be because the speech track referred to 'supplement of histidine'; and the label (7) included the word 'histidine' twice. The shifting of fixations may indicate a search for the 'histidine' referent from (7), (8), (9).

(10) All subjects shifted to the revealed 'incubate' label, and shifted between this label and the culture object and its label. Three subjects refixated among the test tube objects as, leaving the sequence on the 'histidine' object and label.

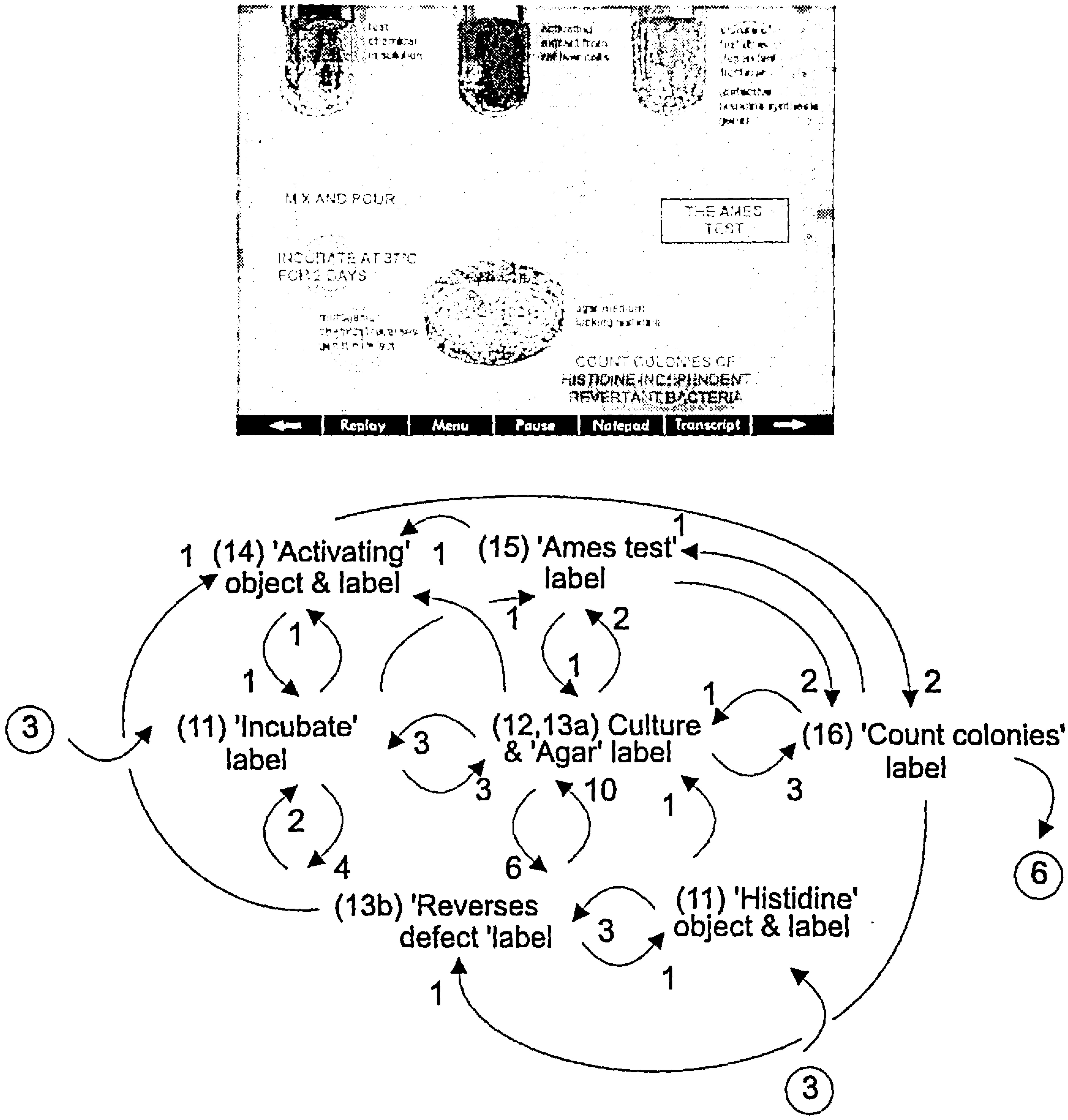


Figure 4.4.7

(11) Fixations were initially spread over the presentation, three subjects in the area of the 'Histidine' test tube and three in the 'Incubate label' and object. Fixations shifted to the 'reverses defect' label when its was revealed.

(12) Fixations shifted to 'culture' object when an animation of the colonies growing began. Fixations were held on the 'culture' object for the short animation.

(13a) (13b) Fixations then shifted between the culture object and 'reverses defect' label, with some subject also returning to the 'incubate' label. From here several subject's fixations again spread out, refixating the 'Ames test' (14) and'Activating label' (15) in particular.

(16) Fixations shifted to the revealed label 'count colonies'.

4.4.3 Discussion

The study provided further evidence for the effect of design techniques upon fixation and attention within presentations. Several conclusions may be drawn :

- Reveal effects had a very strong ordering effect initially in the presentation, with fixations moving from the (2) test chemical and test tube object, to the rate liver cells (3), to the hisitidine dependant bacteria, (4). By gradually revealing labels, symbols and objects, fixations can be shifted.

- Symbols direct attention to other objects. The line symbol and label 'mix and pour' may have shifted fixations downward in the presentation to the agar plate (5).

- Contact points between the speech track and labels or between labels caused the cued object and label to be refixated. This was particularly the case for histidine dependant bacteria, which was cued by 'without a histidine supplement'; and by the label 'agar medium lacking in hisitidine' (6).

- It is important that the object can be identified for the contact point to be made. In (6) the subjects refixated amongst the test tubes (7), (8), (9), probably searching for the referent, which was (7). It is important to emphasise an object with which a contact point is being made, such as using a highlight or bold font.

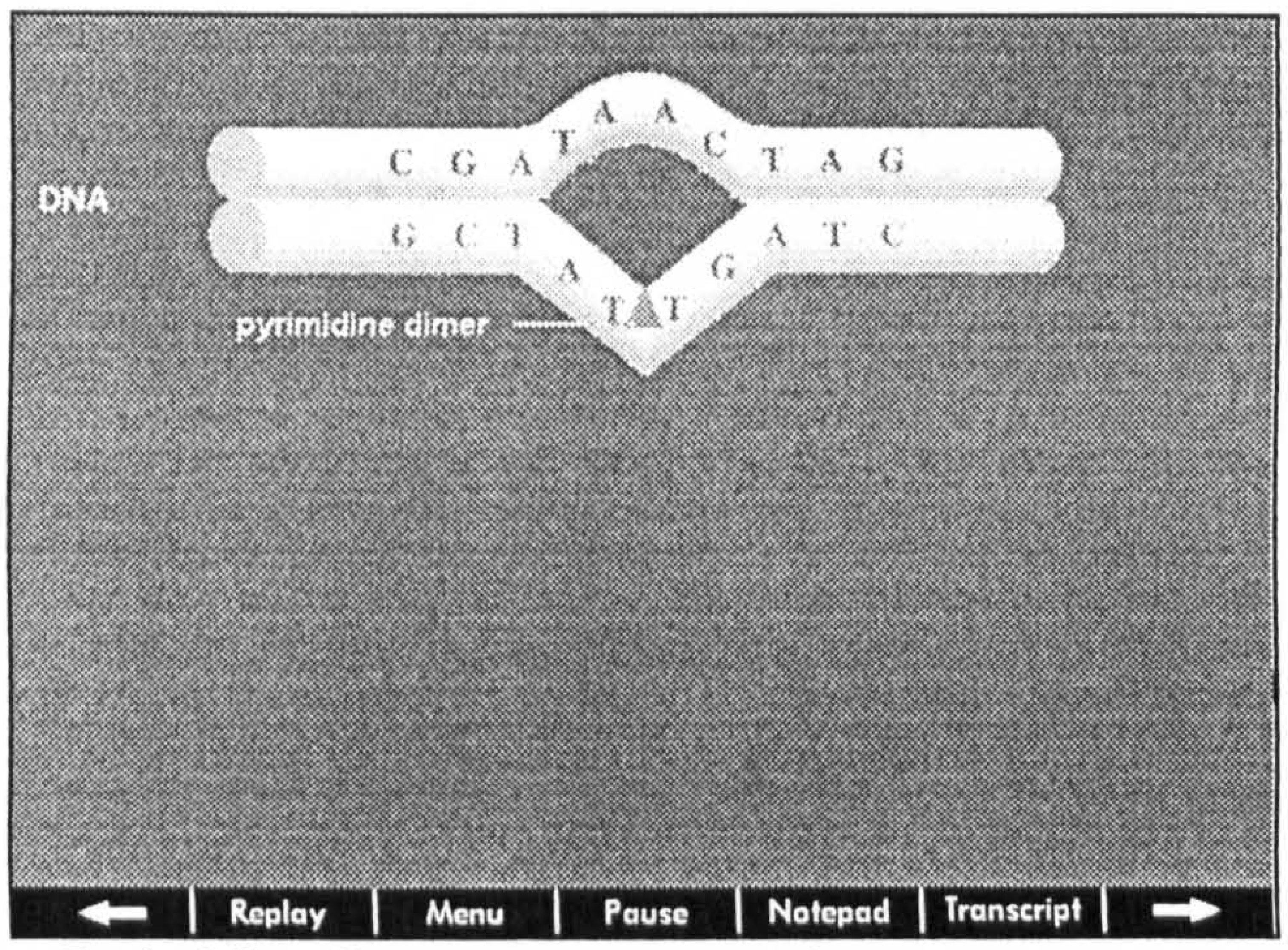
- A large number of objects and labels may cause fixations to be unpredictable. Toward the end of the presentation, the screen was cluttered with labels. Fixations were spread out over the presentation (11). It may be useful to remove certain elements, such as the 'Ames test label' to earlier screens to reduce clutter, and to use emphasise effects to order fixations.

- Animations will attract attention. The appearance of the animation of the bacteria growing on the culture object re-aligned attention to the culture (12)

4.5 'Photoreactivation' sequence

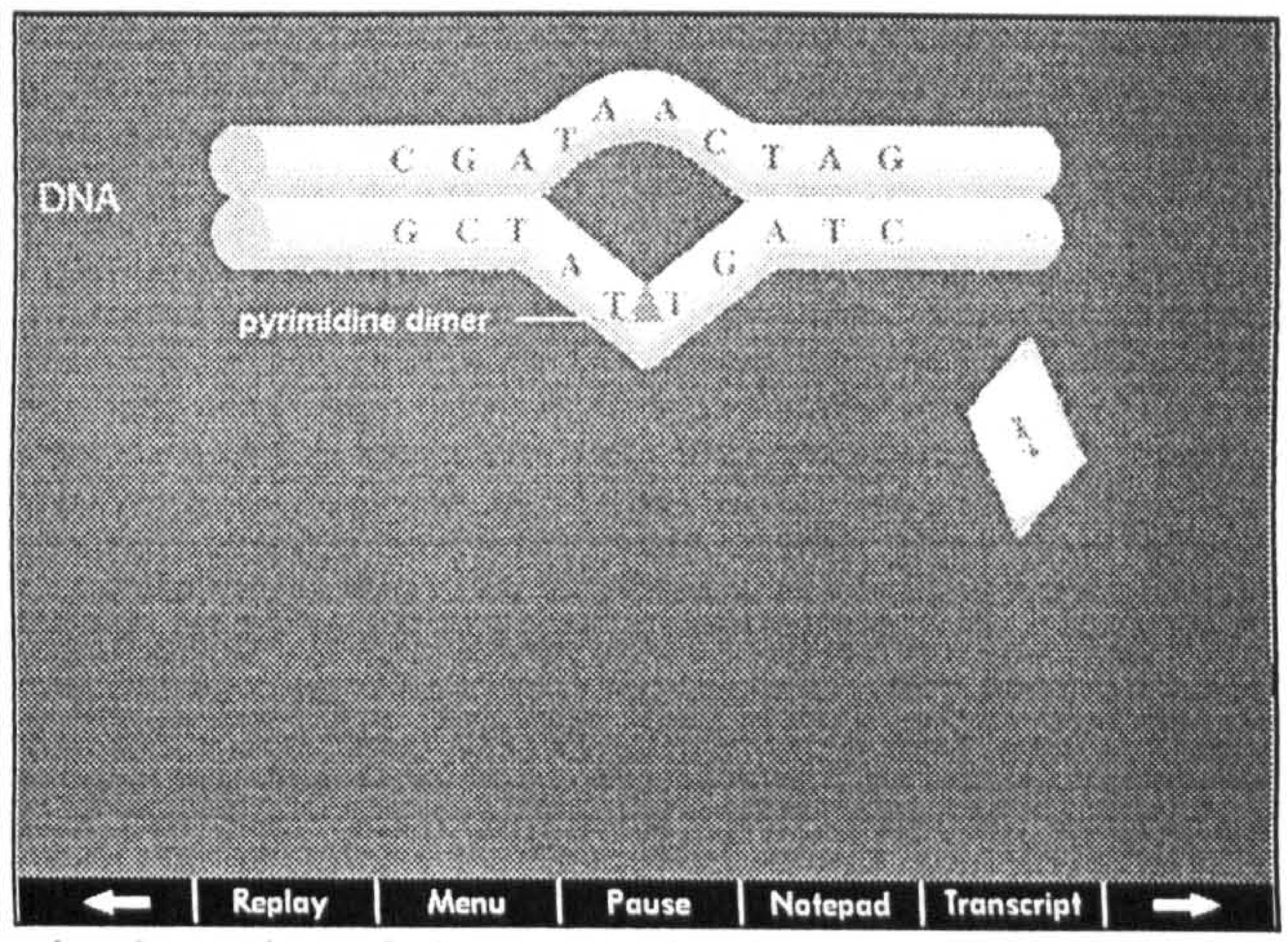
The Photoreactivation sequence shows how DNA damaged by cancer can be repaired. In this case the damage is caused by a dimer, which pushes apart the DNA strands. Repair requires that a special chemical, called a photolyase, binds with the damaged part of the DNA. This creates a protein-dimer complex. A process, termed photoreactivation, then occurs if the complex is struck by light energy. This removes the dimer and reverses the damage to the DNA.

4.5.1 Materials for 'Photoreactivation' sequence



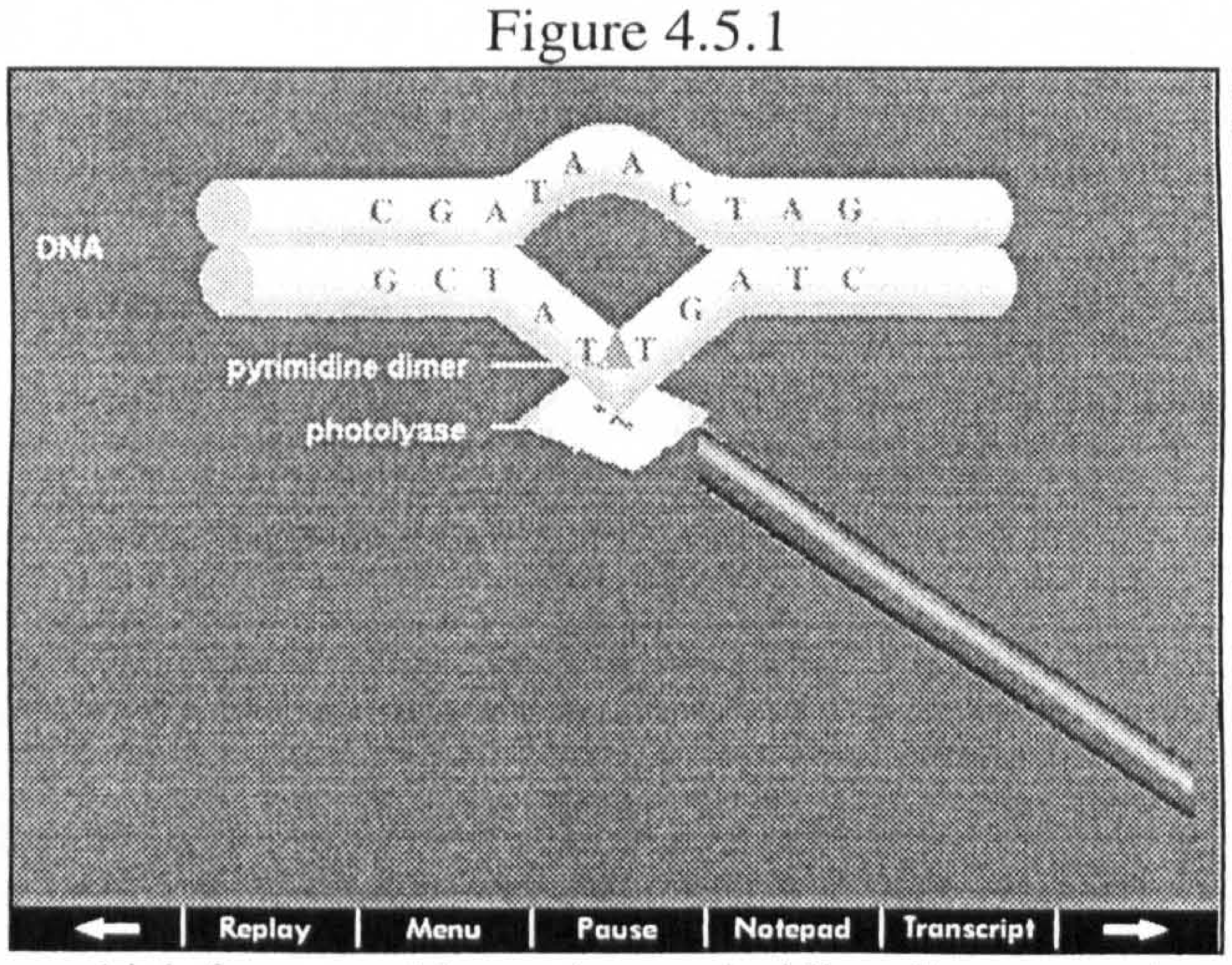
Pyrimidine dimers are reconverted to monomers ...

Figure 4.5.1



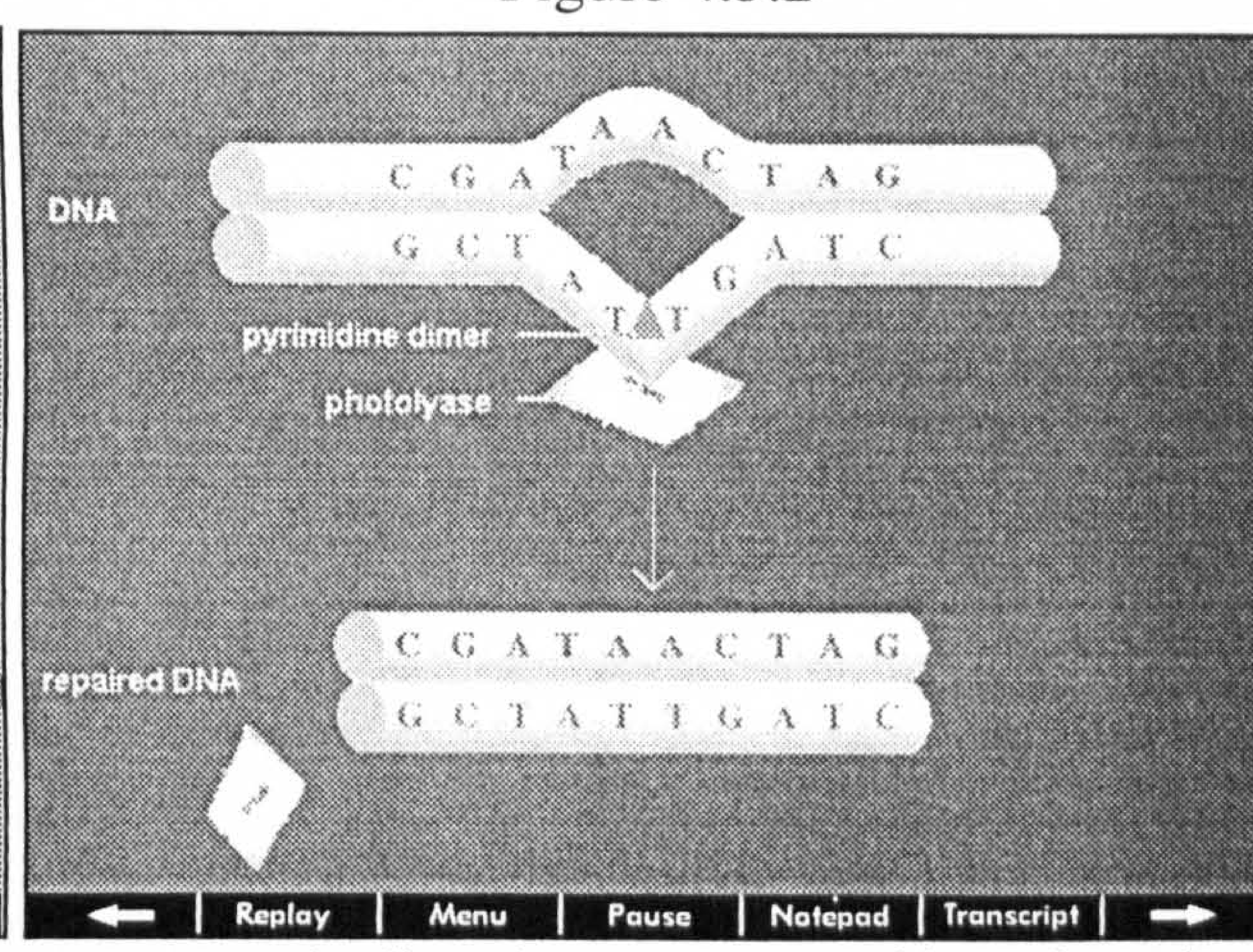
by the action of photo- reactivating enzymes called photolyases ...

Figure 4.5.2



which form a stable protien-pyrimidine dimer complex and then use light energy absorbed from the visible range

Figure 4.5.3



to convert the dimer into monomers without disrupting the double strands.

Figure 4.5.4

Figure 4.5.1 shows the initial state of the presentation, with the DNA and pyrimidine dimer labelled. In figure 4.5.2 the photolyase object is shown moving across the screen from right to left toward and attaching to the pyrimidine dimer. In figure 4.5.3 the photolyase is struck by light energy from the lower right of the screen to the centre of the photolyase. In figure 4.5.4

an arrow, a second photolyase and the repaired DNA is revealed; and then the photolyase object is shown moving off the screen to the left.

4.5.2 Results

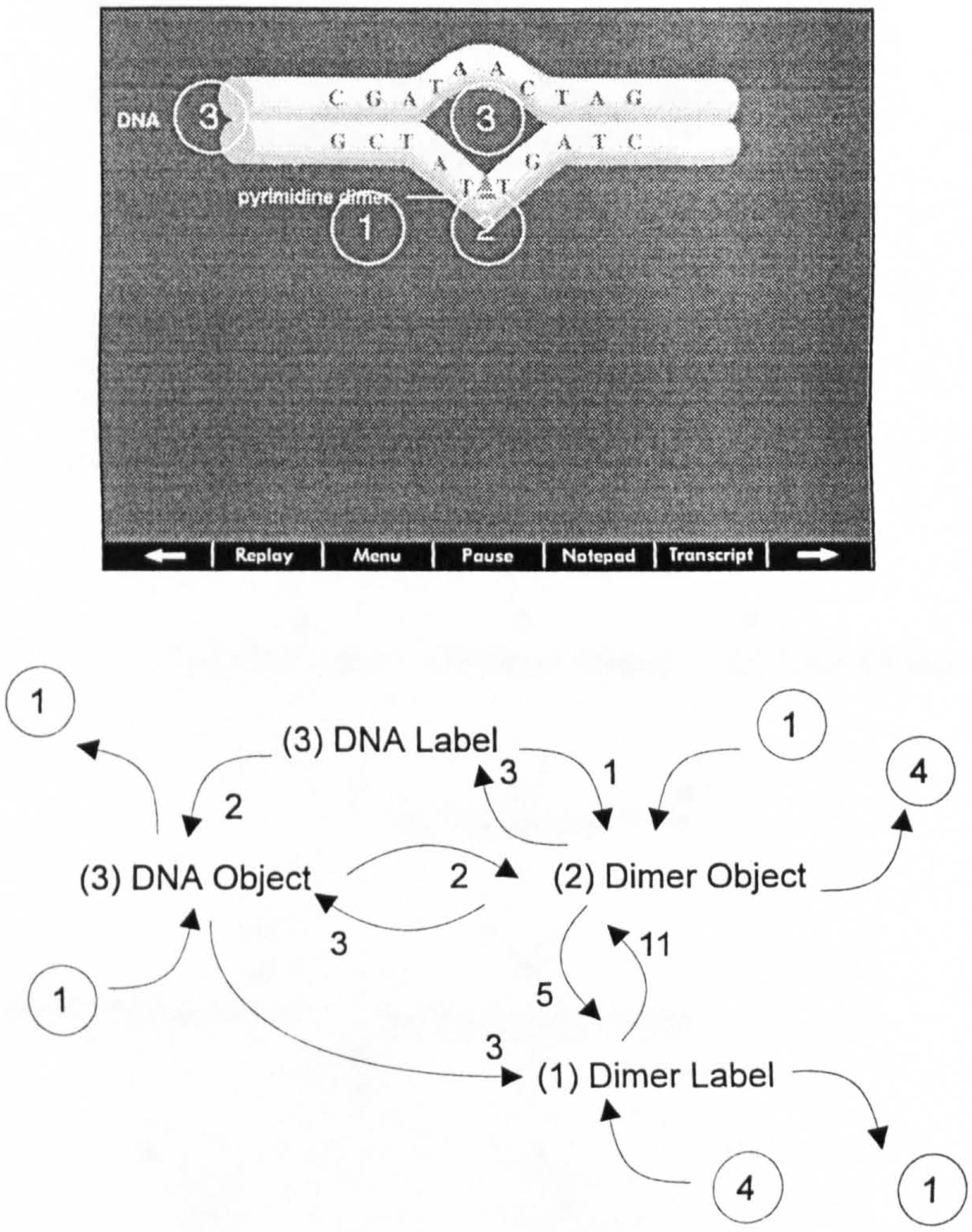
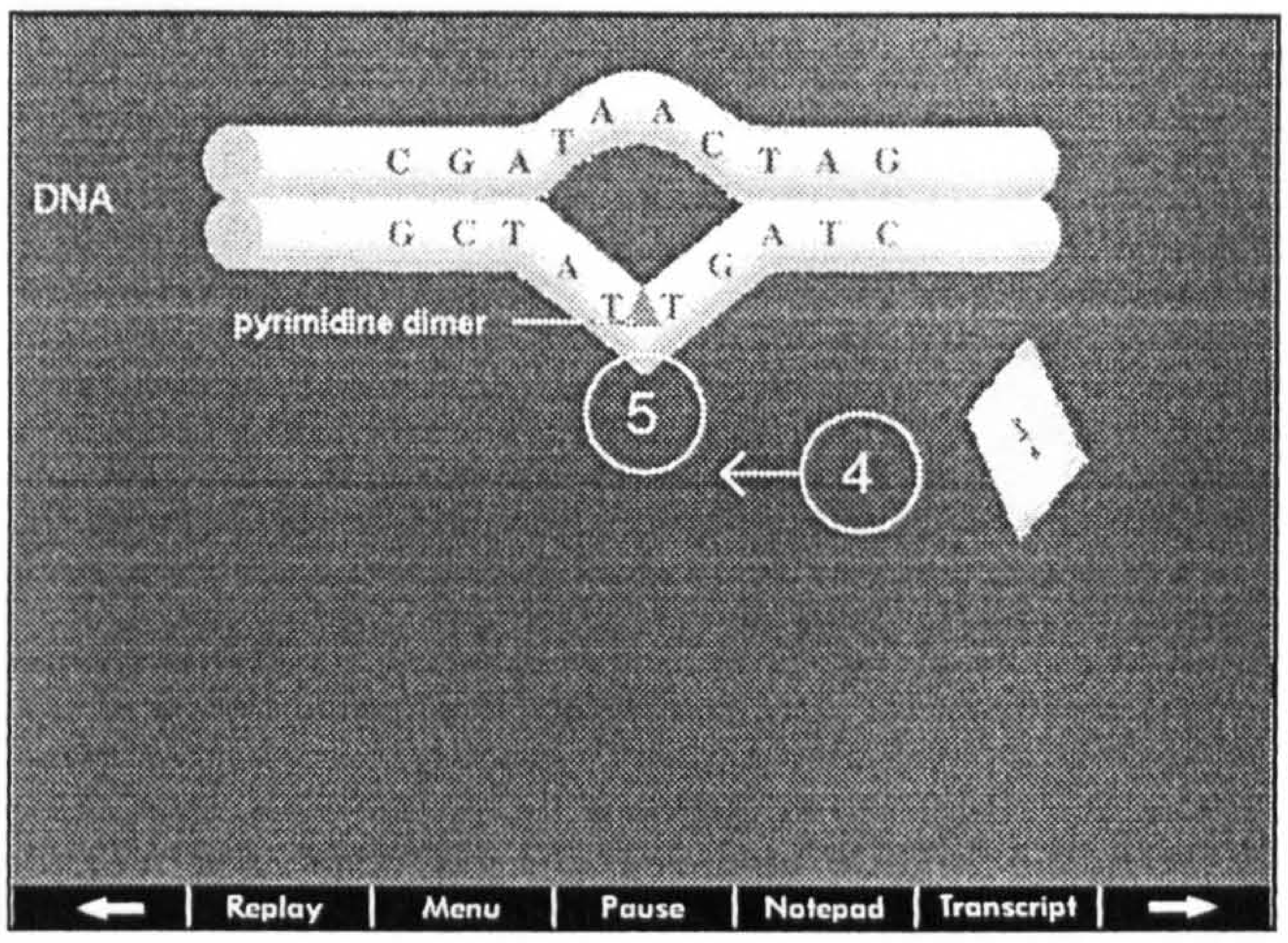


Figure 4.5.5

(1) Five subjects initially fixated on the pyrimidine dimer label and object; one subject fixated on on the deformed, central part of the DNA.

(2) Five subjects shifted to the pyrimidine dimer and switched back and forth to its static label. The pyrimidine dimer object and label received a greater number of fixations than the DNA object and label; possibly because the pyrimidine dimer was cued within the speech track.

(3) Following fixation upon the pyrimidine dimer and label, three of the subjects then shifted between the DNA object and its static label. Fewer fixations may have been given to the DNA label than to the pyrimidine dimer label because unlike (2) it was uncued in the speech track.



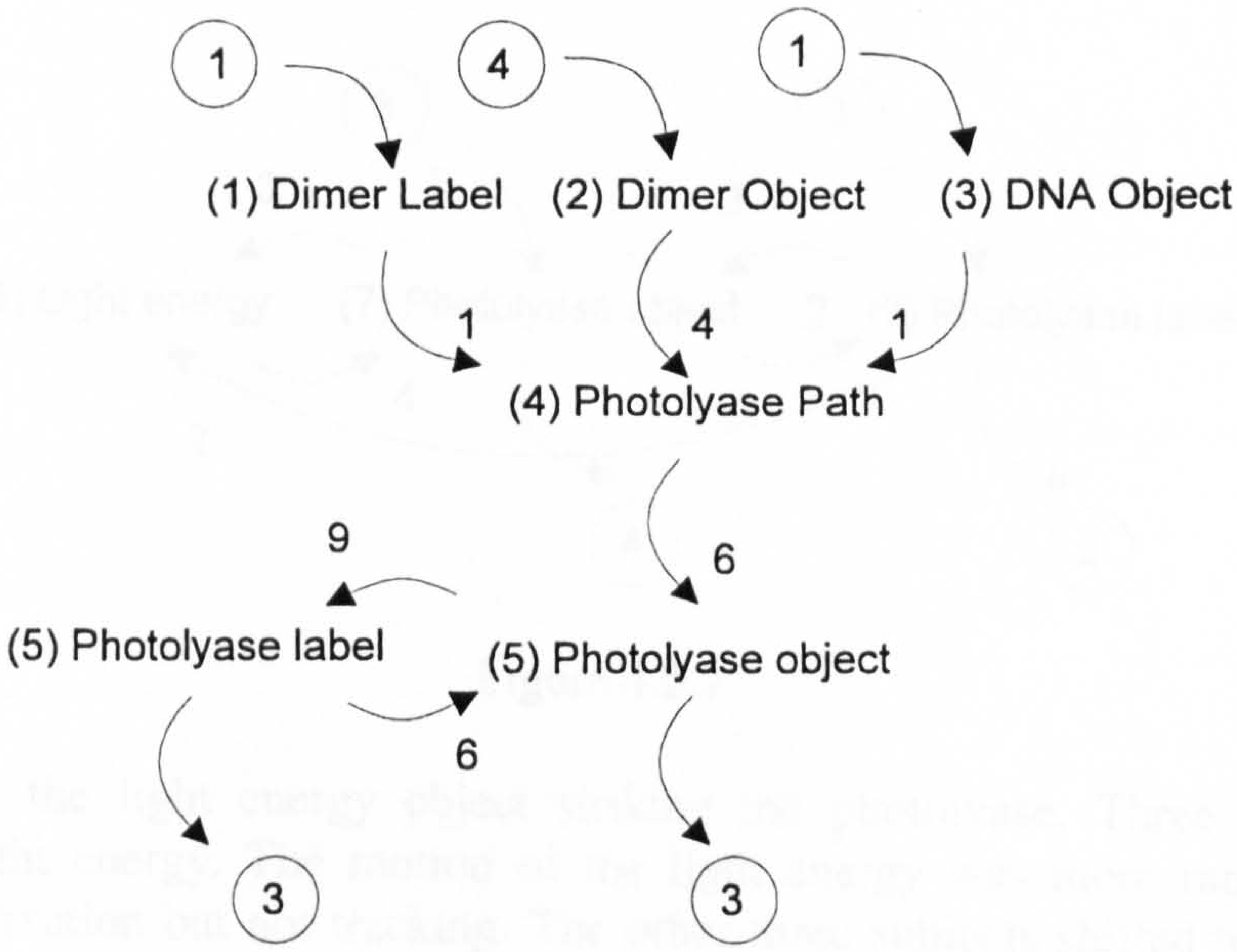


Figure 4.5.6

(4) An animation of the photolyase object appearing and then moving along a path toward the pyrimidine dimer object was shown. All six subjects fixated upon the photolyase object when it appeared, with subsequent fixations tracking its path.

(5) Fixations followed the photolyase object's path to its end point, where revealing the photolyase label caused a shift between it and it's label for all subjects. The photolyase label was also cued by the speech track.

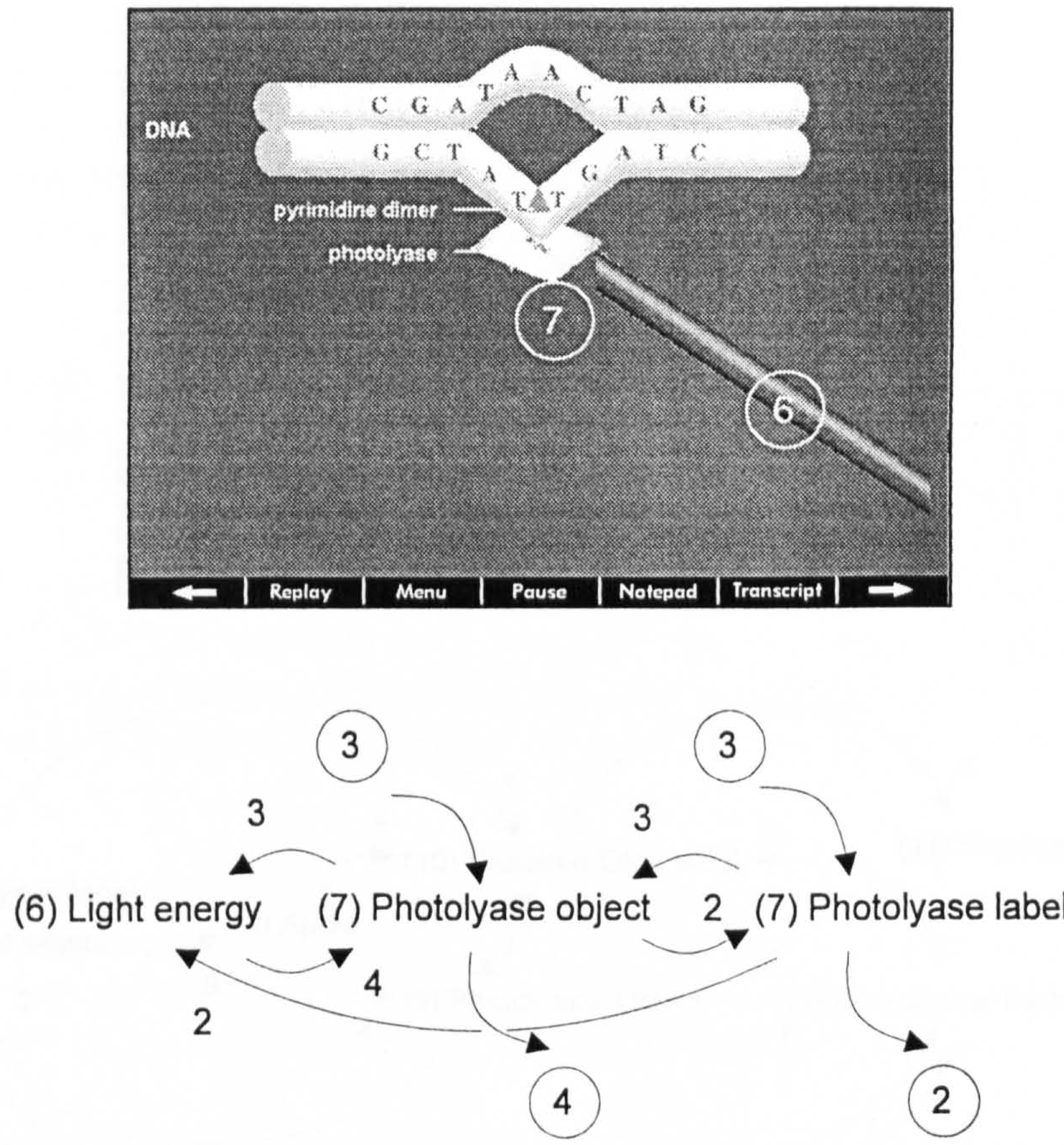


Figure 4.5.7

(6) Animation of the light energy object striking the photolyase. Three subjects shifted fixation to the light energy. The motion of the light energy was more rapid than for (5), causing shifts of fixation but not tracking. The other three subjects shifted to the photolyase object but did not shift further for the light energy motion. This may be because the motion of the light energy was very rapid, giving little time for a shift for those subject previously fixated on the photolyase label.

(7) All subjects shifted back to the photolyase object and label; fixations may have been directed by the end point of the light object's motion in (6). Attention to the pyrimidine dimer may also have been partially cued by the phrase 'protein pyrimidine complex' in the speech track.

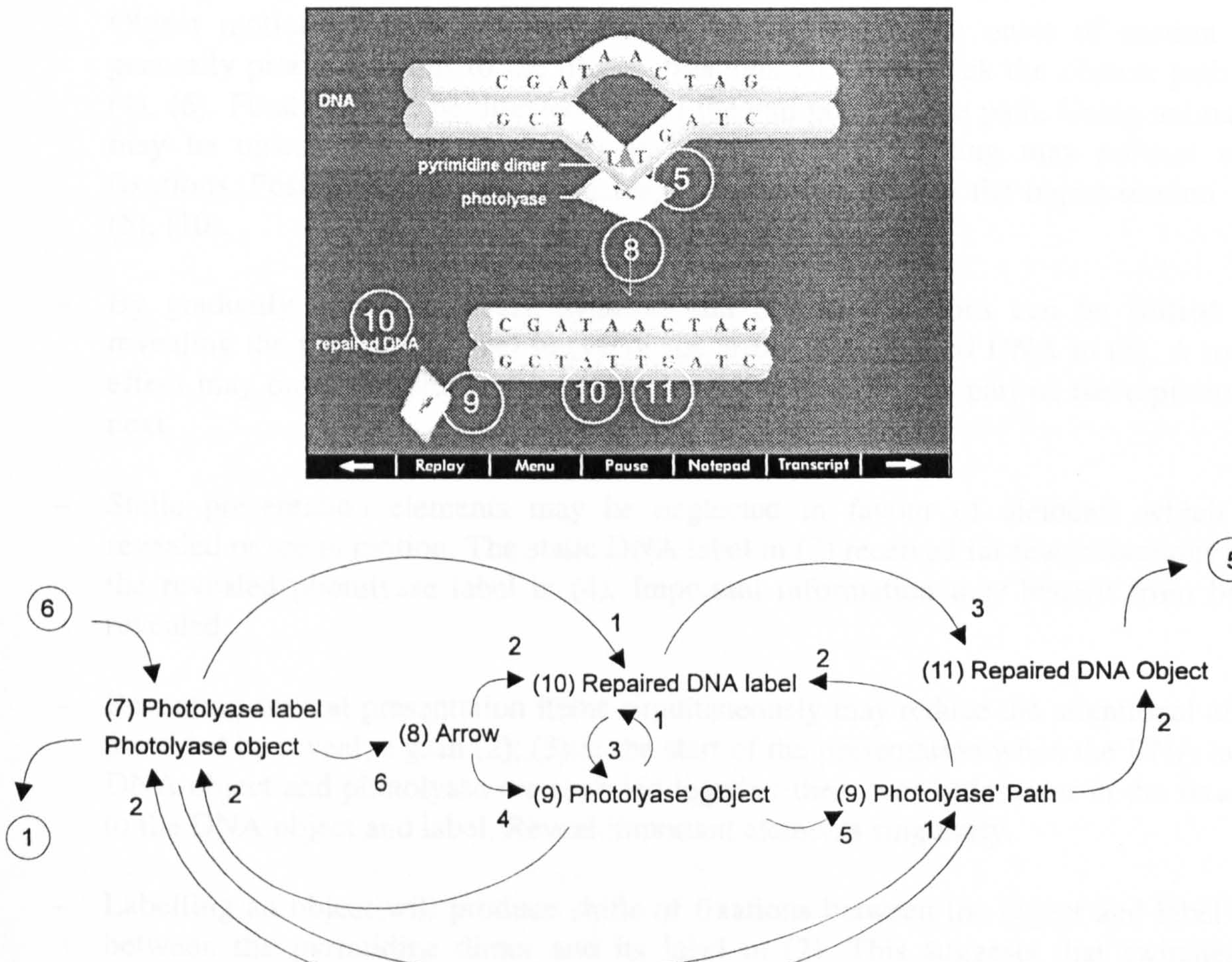


Figure 4.5.7

(8) All subjects shifted from the photolyase to the revealed arrow symbol; fixation followed the arrow down the presentation. Four subjects then fixated on the second photolyase object in the centre of the repaired DNA object; two on the repaired DNA label.

(9) Animation of the second photolyase object detaching and moving away. Five subjects initially tracked the photolyase's path. Two subjects broke off tracking to re-fixate upon the first photolyase object (7); this may have been to compare the two similar shaped photolyase objects.

(10) After (9), two subjects shifted to the repaired DNA label. Tracking the path of the photolyase in (9) may have directed fixation to the repaired DNA label which lies at the end of the path.

(11) Five of the subjects shifted to the repaired DNA object at the end of the presentation. Fixations were generally upon the centre of the repaired DNA from where the second photolyase had departed in (9).

4.5.3 Discussion

- Object motion provides a strong control on attention. The onset of motion will generally produce a shift to the object; fixations will then track the objects path e.g. (4), (6). Fixations will be directed toward the end point of the path. Using animation may be unwarranted if the object is not important; tracking may prevent other fixations. Position subsequent elements at the end of paths of the object motion e.g. (5), (10).

- By gradually revealing labels, symbols and objects, fixations can be shifted e.g. revealing the photolyase label in (v), or the arrow and repaired DNA in (8). A reveal effect may order the processing of the presentation from one part of the topic to the next.

- Static presentation elements may be neglected in favour of elements which are revealed or are in motion. The static DNA label in (3) received far fewer fixations than the revealed photolyase label in (4). Important information may benefit from being revealed.

- Revealing several presentation items simultaneously may reduce the attentional effect provided by reveal; e.g. in (2), (3) at the start of the presentation when the DNA label, DNA object and photolyase are revealed together there was little order in the fixation to the DNA object and label. Reveal important elements singularly.

- Labelling an object will produce shifts of fixations between the object and label e.g. between the pyrimidine dimer and its label in (2). This suggests that switches of attention are used to integrate information from the label with the object. Ensuring that an object and its label are both appear at the same time may improve this process.

- Cueing of labels within the speech track will produce a shift of fixations to the object and it's label. The pyrimidine dimer (2) and photolyase (5) were both re-fixated when cued. Speech cues should be tightly synchronised to labels. This may be particularly important in allowing the speech track to be integrated with the object and its label : reveal objects and labels when cued in the speech track.

- Symbols may direct fixations to other objects, rather than upon the symbol itself. The arrow symbol (8) shifted fixations to the repaired DNA object. The learnt meaning of symbols may be used to control or guide fixation within the presentation.

4.6 Conclusions

The aim of this chapter was to investigate the issue of attentional design. The studies reported have shown some of the dangers of attempting to design multimedia presentations without concern for attention and comprehension. Multimedia designers should consider how to control shifts of attention according to the content being shown.

Attention and the viewing sequence can be influenced by design of highlighting, and reinforcement with speech. Designers also need to model the user's viewing process to provide time for reading captions and to avoid attentional clashes. If the presentation is complex it may also be useful to guide attention to important elements using reveals, labels, symbols or speech. Attentional design may be more important for low domain knowledge subjects, who need to be guided at exactly what to attend to.

Animation

- Animation has a strong effect on attention. The onset of an animation will generally produce a shift of attention to the object in motion if attention is not already engaged elsewhere in the presentation. If the start of the motion is important, then ensure that attention will be free, not processing labels or captions.

- Attention will generally be locked when processing motion, making it unavailable to shift to other elements which change or are revealed. Using animation may be unwarranted if the object is not important; tracking may prevent other fixations

- Animation which shows an object moving as a whole along a path produces fixations which will track the motion of the object and will be directed toward the end point of the path.

- Animation which shows an object changing state, in which the object remains as a whole but it shape alters, produces dispersed fixations over the area and edges of the object.

- Changing scenes causes attention to re-orientate. Each time attention seemed to process the scene from fresh, e.g. refixating on the previously attended title after each scene change. Avoid revealing labels or object motion to allow re-orientation to take place after changing to a new scene.

Text

- Labelling an object will produce shifts of fixations between the object and label. Ensuring that an object and its label are both available together may aid this process.

- Reading captions and labels will lock attention, preventing other elements being processed. Reading also requires longer fixation time than for identifying objects. Allow time for text elements to be processed.

Still Image

- Static object elements are not attended to uniformly. Those which are labelled, or appear to be different from their surroundings will be attended to in preference to other objects.

- Reveal effects had a very strong ordering effect initially in the presentation. By gradually revealing labels, symbols and objects, fixations can be shifted.

- Symbols may direct fixations to other objects, rather than upon the symbol itself. The learnt meaning of symbols may be used to control or guide fixation within the presentation.

- Labelling an object will produce shifts of fixations between the object and label. Ensuring that an object and its label are both appear at the same time may improve this process.

- Revealing several presentation items simultaneously may reduce the attentional effect provided by reveal. Reveal important elements singularly.

Contact Points

- It is important that the object can be identified for the contact point to be made. It is useful to emphasise an object with which a contact point is being made. Cueing of labels within the speech track will produce a shift of fixations to the object and it's label. This may be particularly important in allowing the speech track to be integrated with the object and its label : reveal objects and labels when cued in the speech track.

- A large number of objects and labels may cause fixations to be unpredictable. It may be useful to remove certain elements to earlier screens to reduce clutter, and to use emphasise effects to order fixations.

It may be dangerous to over generalise the results of the studies for several reasons. First, the studies made use of a relatively small sample of subjects, all of whom were from a common background. The study did not manipulate domain knowledge, which is likely to have some impact upon viewing sequence.

Second, the presentations used were all strongly driven by narrative and had for the most part a clear viewing sequence. For presentations with a less clear structure it may prove more difficult to provide such strong guidelines. In such cases, it would also be expected that individual differences would be far more strongly evidenced in the results.

Third, the studies were exploratory, in that they used existing materials which were not specifically designed to test any particular guideline. This will lead to some doubt as to whether any guideline has a particular effect, or whether the effect is due to a combination of factors. Because of the number of possible guidelines combinations, and the difficulties in conducting the studies and performing the data analysis, the studies only used a single experimental condition. This makes the results weak, since they do not provide 'with problem - without problem' conditions to test if the guidelines actually predict an attentional problem.

Chapter 5.

Study : Comprehension of the Etiology of Cancer

5.1 Introduction

The second set of key issues which the framework for multimedia design suggested were within media design, and media selection.

This chapter presents a study which re-authored the multimedia presentations shown in chapter 4 to investigate the impact of within media design. It also compared the effectiveness of language against language and visual media in presenting the content. The value of the design changes was measured by testing the subject's recall of the content.

Because of the number of possible guidelines combinations, the studies could not exhaustively test the role of each guideline in comprehension. Instead, the study focuses upon whether the guidelines can solve design problems, and thus improve subject's recall.

5.2 Method

The study had three conditions, summarised in table 1. The first and second condition used the original multimedia presentations from the eye tracking studies. These conditions used two groups of subjects. Condition 1 took a group drawn from research students within City University, who had low domain knowledge. The condition 2 group were biochemistry post-graduates who had a high general domain knowledge of anatomy and physiology but no detailed knowledge of the process of cancer invasion. The subjects were all pretested to ensure that they had no specific knowledge of the subject matter itself. Both groups were shown the presentation once and were then instructed to write down all that they could recall of the presentation. The subjects were told that spelling was not important, but that they should try to be clear as to what happened, and in what order.

Condition Number	Materials Used	Subjects
Condition 1	Original MM materials	Novice
Condition 2		Expert
Condition 3	Re-authored MM materials	Novice
Condition 4	Text & Speech only	Novice

Table 1 Conditions used in recall studies

The third condition used a set of re-authored materials, based upon the method's guidelines. This was to investigate whether the design guidelines would improve recall. For this condition, a further set of subjects were recruited from research students within City University, who again had low domain knowledge. High domain subjects were not used in this condition because it proved impractical to recruit further subjects with a similar background to those in the first study. The subjects were again shown the presentation once only, and then scored for recall in the same way as before.

The fourth condition used text and speech only materials. This was done to be sure of how re-authoring had effected recall. Again, low domain subjects were used. The text script derived from the presentation was shown, and was accompanied by a matching speech track. This condition acted as a control condition to compare the original and re-authored design. It addressed a particular concern in re-authoring that more text and speech had been added e.g. in describing the motion of an object which was previously shown only as animation. If recall was improved above that found for the original presentation, then this could be attributed to the addition of text captions and the changes in the speech track. If the recall was poorer, then the use of the re-authored images and animations had aided performance.

5.2.1 Materials

The study investigated comprehension of the same multimedia presentations which were previously studied using eye tracking, shown in chapter 4.

Re-authored versions of the presentation were produced using Macromedia Director. Care was taken to preserve the same content as the original presentation. New media and presentation techniques were added when the guidelines suggested a failing in the original design. The re-authored presentations were then validated to ensure that they contained the same set of reference propositions as the original presentation.

Whilst the aim of the study was to test out as many guidelines as possible, changes in design were only made when the guidelines suggested that a problem might have arisen in the original presentation. Because of this, the guidelines which were applied tended to concern within media design and attentional design, rather than media selection.

5.2.2 Data Analysis

In order to score the accuracy of subject recall, the multimedia presentations were analysed into a set of reference propositions by an independent expert. The propositions were grouped into either high level organising concepts of procedures, problems or cause-effects, or into lower level actions or objects within each part of the sequence. The subjects written recall was split into propositions, and then matched to the reference propositions. Spelling errors were ignored.

5.3 'Metastasis' sequence

5.3.1 Materials for 'Metastasis' sequence

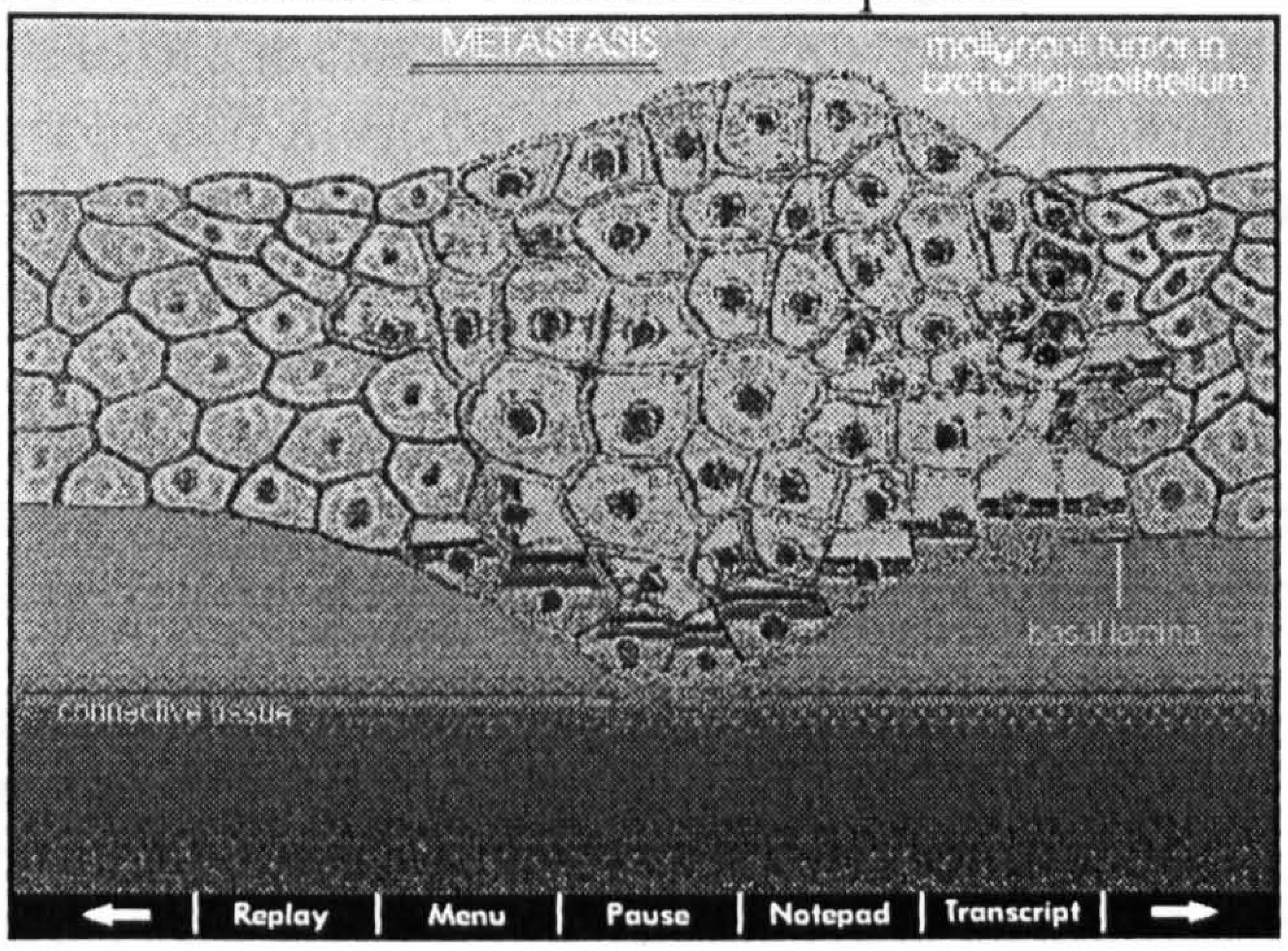


Figure 5.3.1

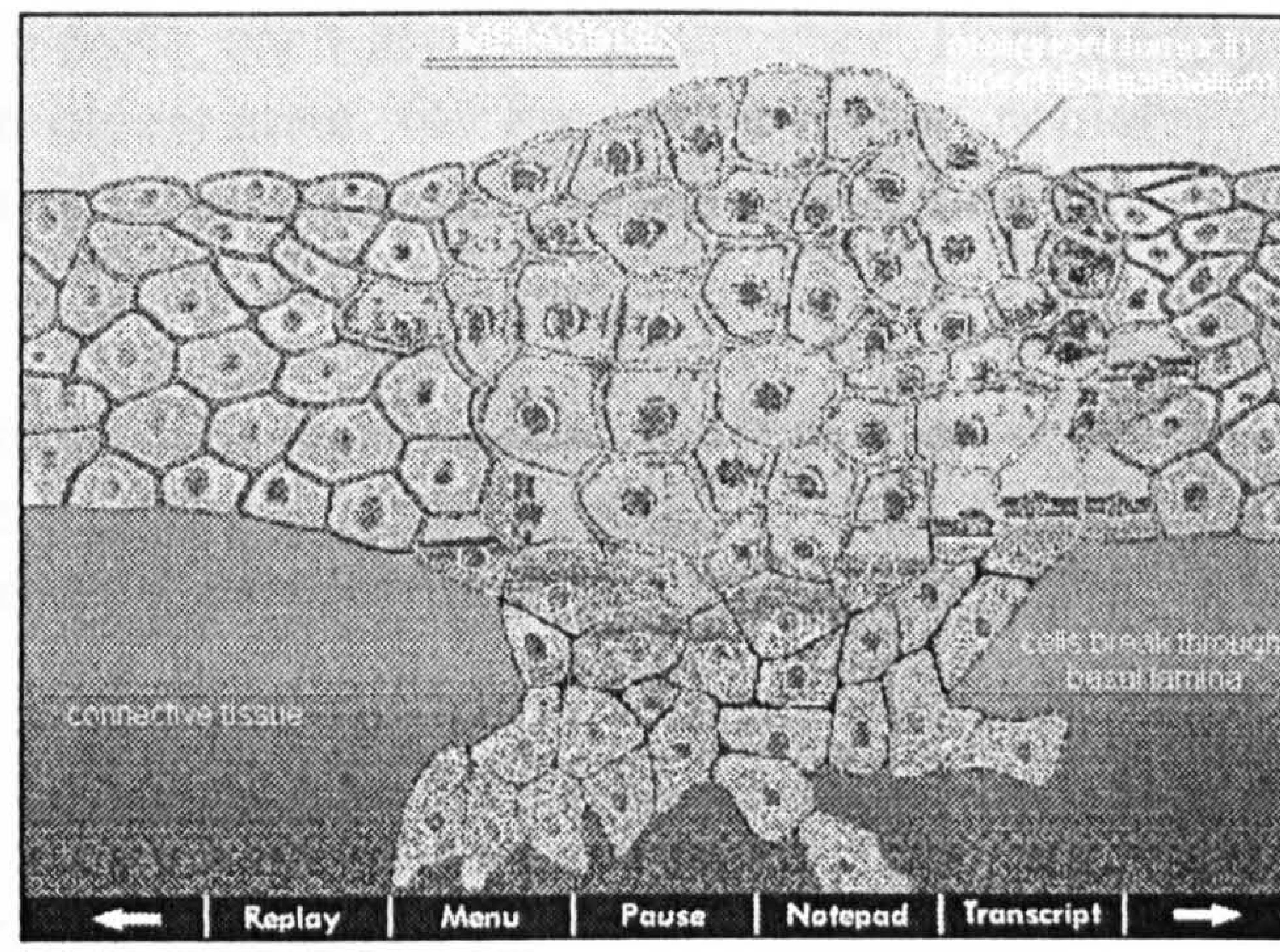
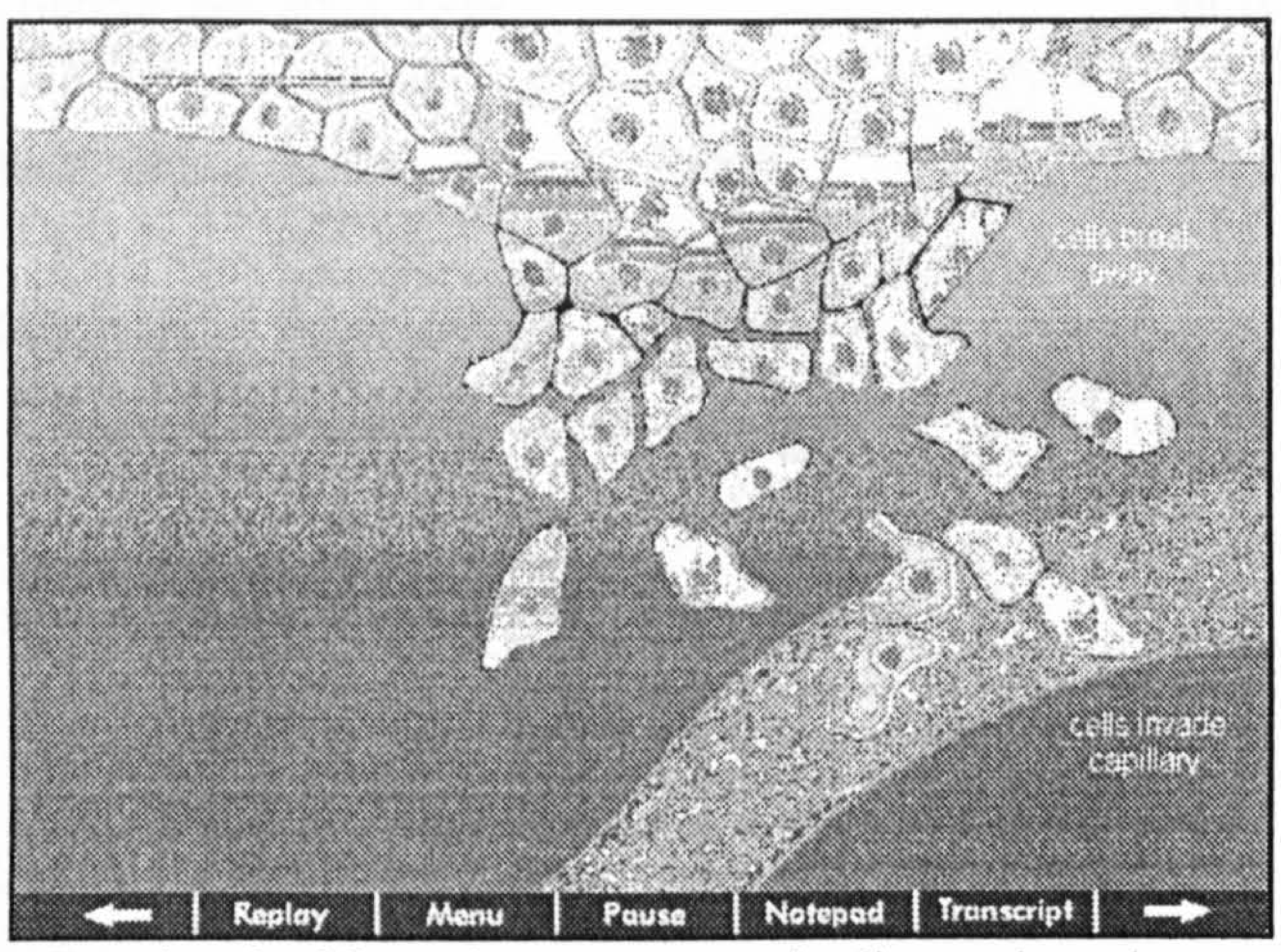
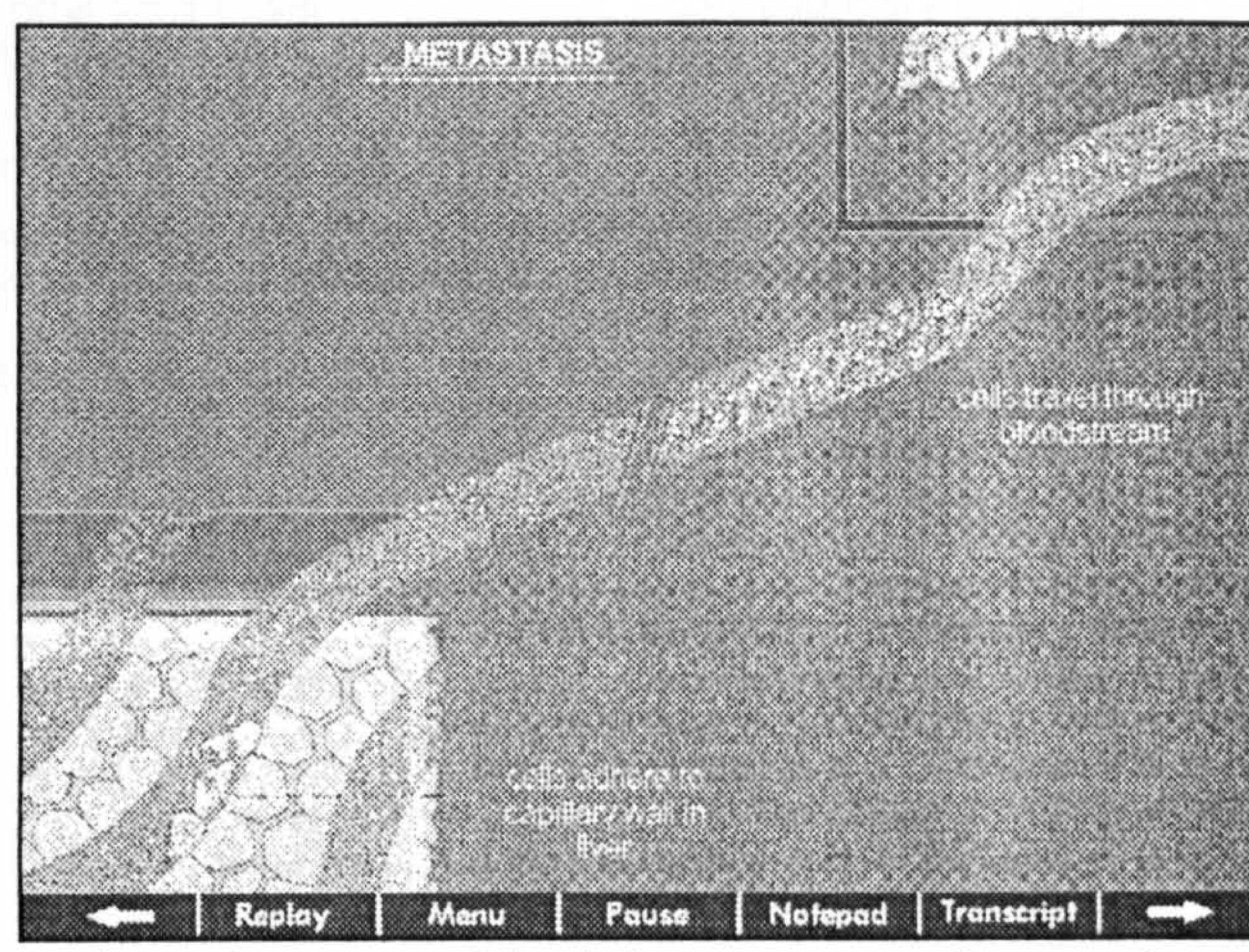


Figure 5.3.2



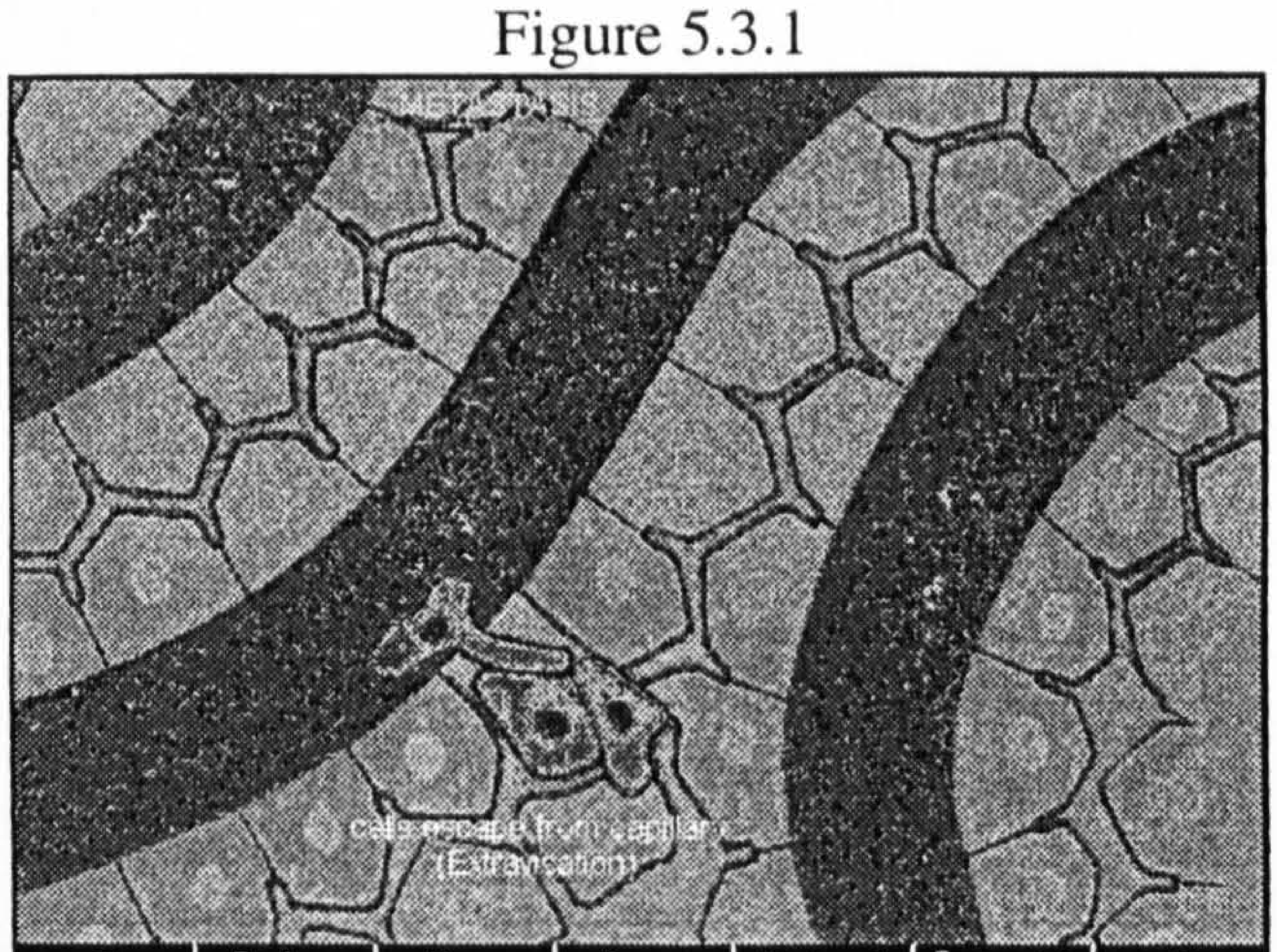
Despite significant improvements in diagnosis and treatment ...

Figure 5.3.1



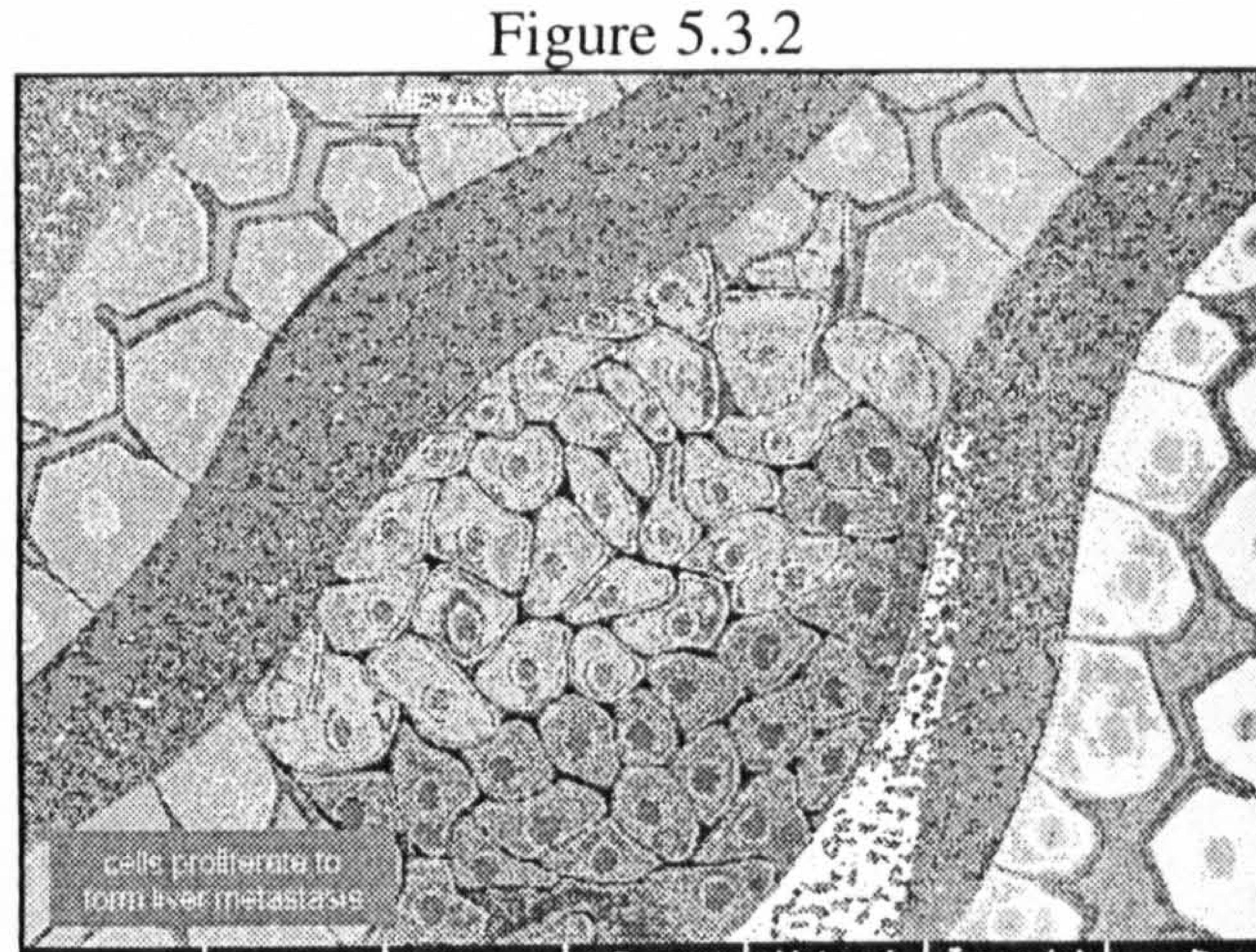
... the major cause of death in most cases of cancer is metastasis ...

Figure 5.3.2



... the spread of tumour cells from the primary neoplasm to distant organs ...

Figure 5.3.3



... where secondary growth occurs.

Figure 5.3.4

For analysis of recall, the metastasis presentation was divided into a set of eleven key propositions (figure 5.3.1). At the top level is the cause of the sequence, that metastasis is the spread of the cancer. This is followed by three main parts, each given as a sequence. First the tumour cell must break away. This begins with an object state, the tumour in the bronchial epithelium. The tumour the breaks through the basal lamina and breaks into the connective tissues.

Next, the tumour cell travels through the blood stream. The tumour cells must first enter the blood vessel capillary, and then adhere elsewhere to the capillary. Finally the cells form a secondary tumour. This requires the cells escape from the capillary and invade another organ.

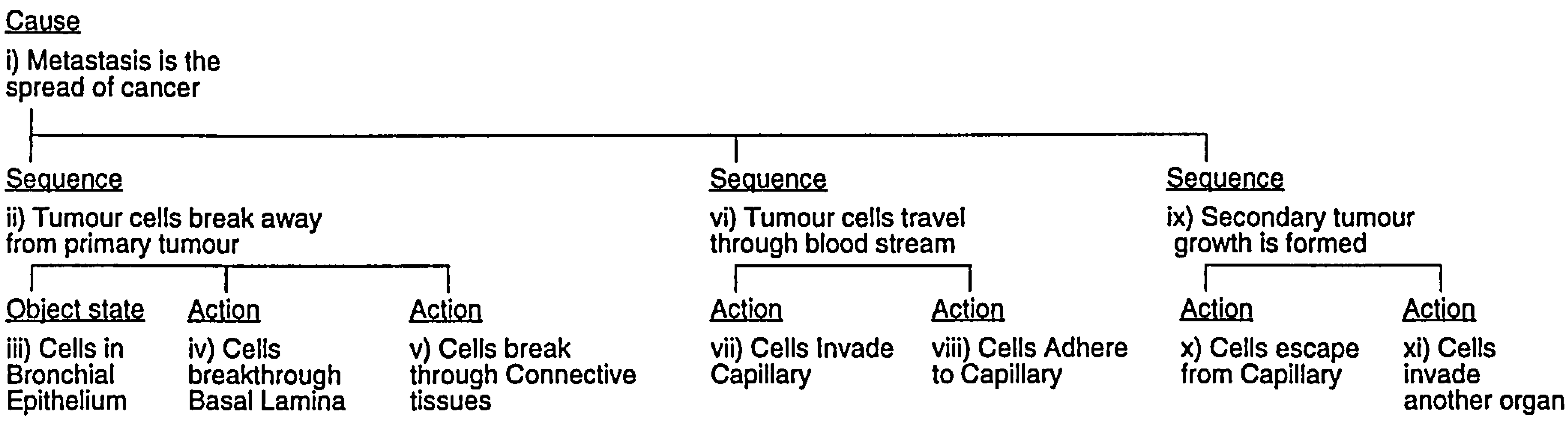


Figure 5.3.1

5.3.2 Results for 'Metastasis' sequence Condition 1 & 2

Proposition type	Proposition	Condition 1 Low Domain knowledge (/8)	Condition 2 High Domain knowledge (/8)	Total (/16)
Cause	i) Metastasis is the spread of cancer.	6	7	13
Sequence	ii) Tumour cells break away from primary tumour.	6	4	10
Object state	iii) Malignant tumour cells are in bronchial epithelium	1	1	2
Action	iv) Tumour cells breakthrough basal lamina	0	2	2
Action	v) Tumour cells break away through the connective tissue	0	1	1
Sequence	vi) Tumour cells travel through the blood stream.	7	5	12
Action	vii) Tumour cells invade the blood vessel capillary	4	5	9
Action	viii) Tumour cells adhere to the capillary wall in liver	0	2	2
Action	ix) Tumour cells escape from capillary	2	4	6
Action	x) Tumour cells invade another organ.	6	5	11
Sequence	xi) A secondary tumour is formed.	8	5	13
	Total :	40	42	

Table 5.3.5

5.3.3 Discussion

The results are shown in table 5.3.1. There was no significant difference in recall between high and low domain knowledge subjects. The background knowledge of the subjects appeared to make little difference to their recall performance. This suggests that performance was based on what was given in the presentation, rather than what was already known by subjects from their existing long term memory.

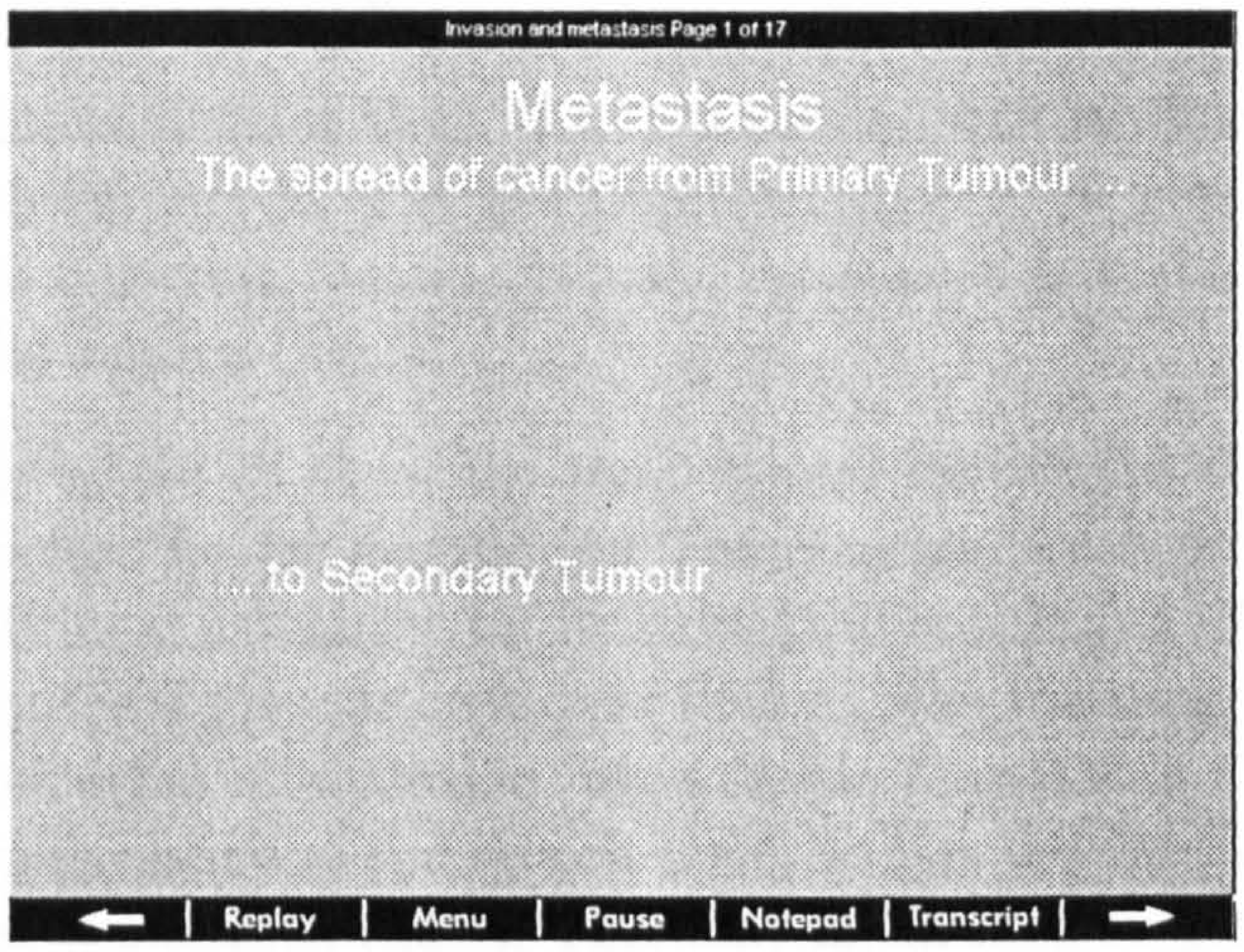
- Recall of higher level sequence propositions ii), vi), xi), representing a group of actions, was good for both expert and novice groups. Most of these sequence propositions were given as animations. Animations are known to improve comprehension of general action information. The use of extra animations may have improved recall performance for e.g. the tumour breaking through the basal lamina, tumour adhere and extravisation actions.

- Recall of specific action, object state and object role propositions iv), v), viii), ix) was poor. Most of these propositions were given in text caption or label elements. Comprehension difficulties may be in part due to attentional problems with the use of multiple labels and animations, discussed earlier. Others problems may be related to poor synchronisation of the text labels with the speech track, see below.

- Linking of the speech track to visual elements of the presentation was generally poor. The speech track rarely primed or re-inforced the propositions presented visually. Where speech did prime the animation toward the end of the presentation in propositions x) and xi) describing the secondary tumour, it improved comprehension. Mostly however, the speech propositions i), xii) did not directly relate to the visual elements being shown. It is important that speech and text captions, labels, animation or objects combine to form a single composite proposition. Comprehension may fail if different propositions are presented by the two media.

- The use of a the general title 'Metastasis' improved recall for proposition i). Titles, advanced organisers and summaries may help to structure comprehension of the presentation.

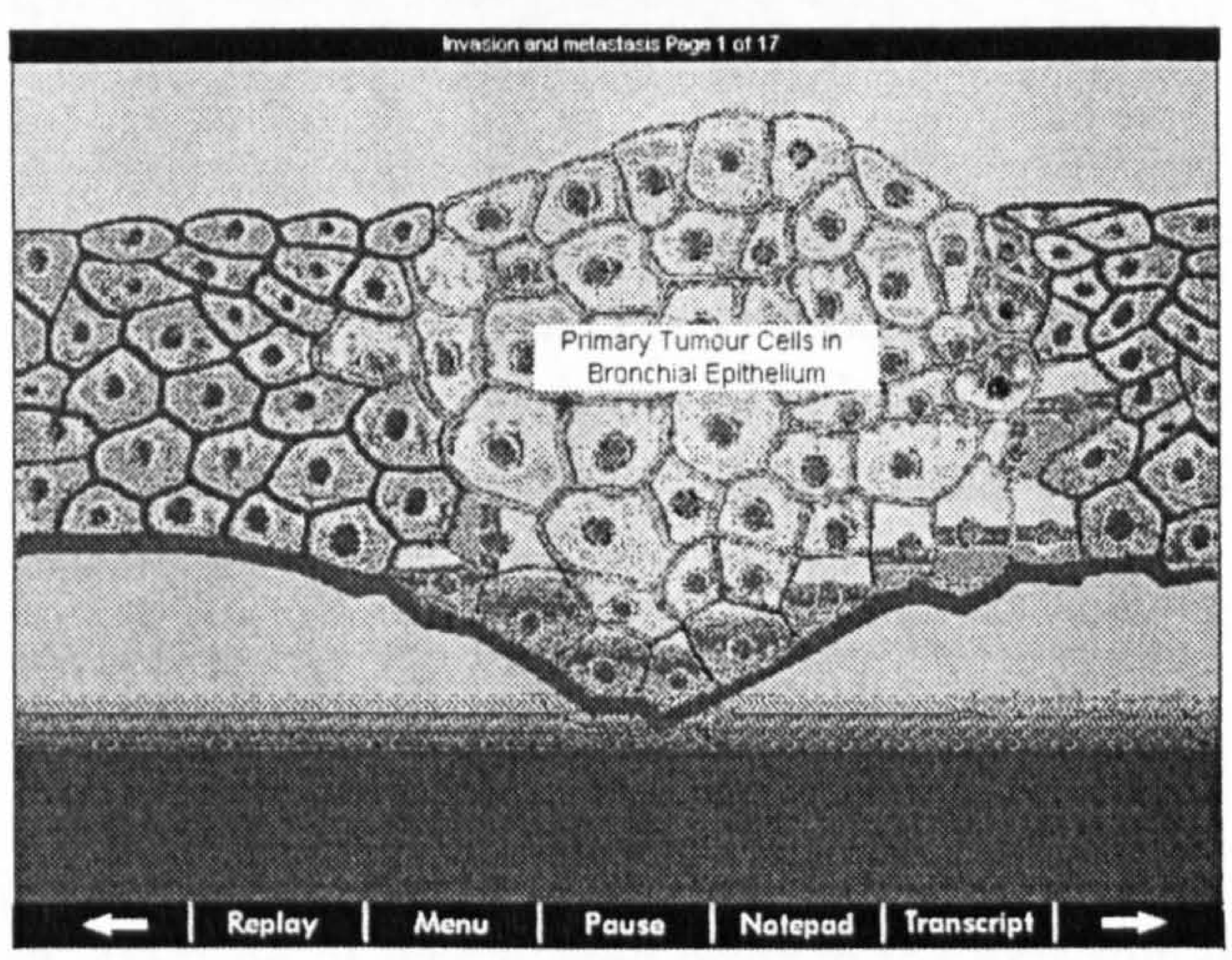
5.3.4 Re-authoring 'Metastasis' sequence



Metastasis is the spread of cancer from primary to secondary tumours.

Figure 5.3.6

The redesign began by providing an advanced organiser for the sequence (LP2), using a caption and speech track which introduced the main topic of Metastasis (fig 5.3.6). The caption was tightly linked to the speech track to reinforce the propositions given (VR5). The organiser will form the top level causal theme, below which other propositions could then be integrated.



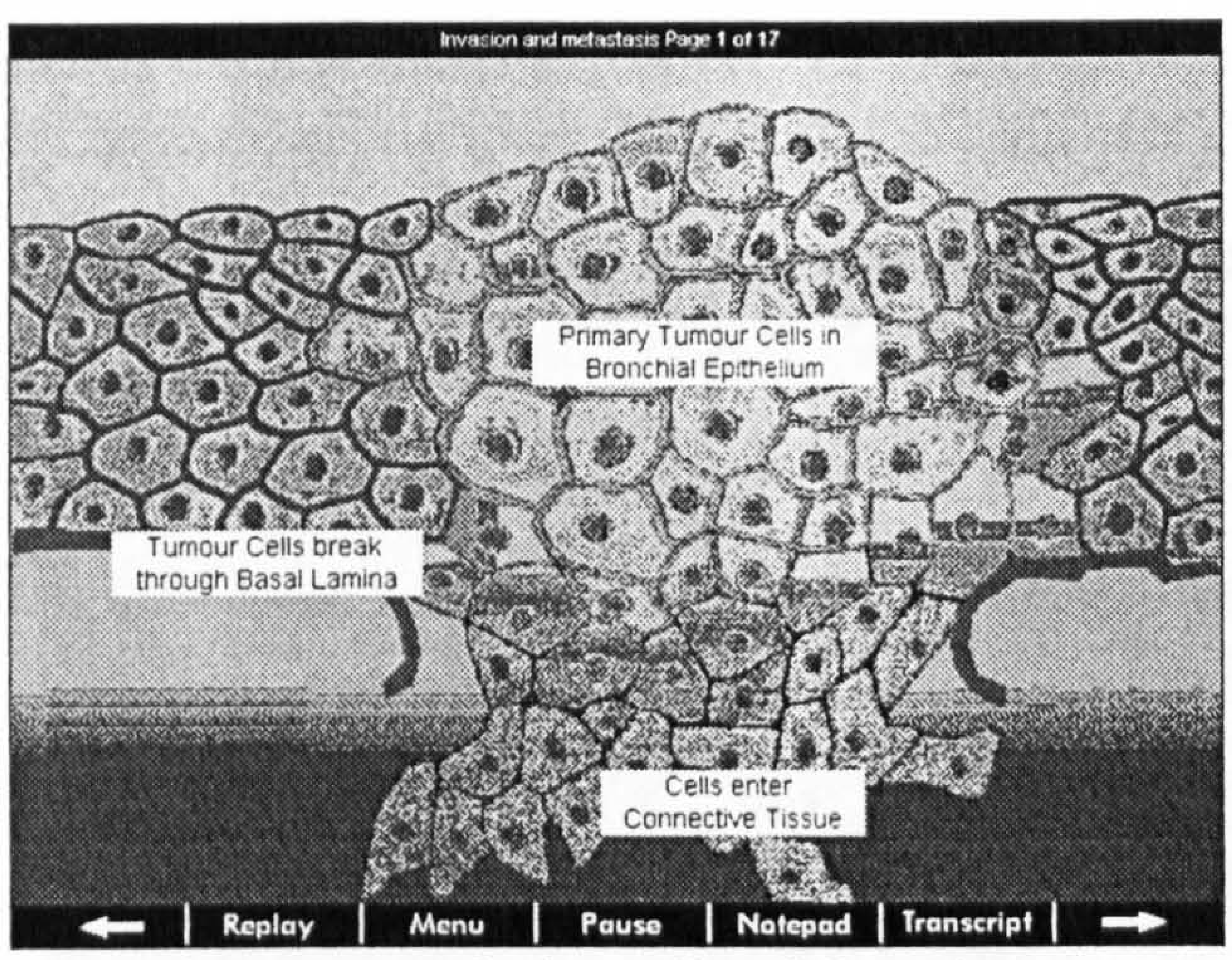
The malignant primary tumour shown is in the patients bronchial epithelium.

Figure 5.3.7

In the first scene itself, the title 'Metastasis' was removed and placed in the organiser as any unnecessary titles may distract attention following orientation (fig 5.3.7). The original presentation also used three other labels, which were shown concurrently. These may cause attention to shift unpredictably between objects (IA7). Instead a single central label was used to direct attention to the central tumour mass (IA5); other labels were then revealed as required. The tumour mass had a different colour to its background, this was left as-is, since objects which stand out against their ground will usually attract attention (IP4).

Following a pause to allow reading time (TA2), a caption was revealed to shift attention to the basal lamina. The basal lamina was not brightly coloured, so a new basal lamina object was drawn which stood out more clearly, to make it easier to attend to (IP4). Again reading time was allowed for the caption (TA2).

Next an animation was added to show the basal lamina split apart (fig 5.3.8). To improve comprehension of the action for the proposition 'Tumour cells break through basal lamina' the next frame used an additional animation of the basal lamina splitting (SR5). The speech track was also changed to reinforce the text captions (VR5). Using the same phrases in speech and captions will tightly link what is processed into a single proposition (VR9). The start of the animation was then synchronised to the speech track, so that the spread of the tumour began as it was introduced by the speech (CP3). This may again help to produce a single unified proposition.



In order to spread, the cells of the primary tumour must break through the basal lamina and then enter the surrounding connective tissues.

Figure 5.3.8

The animation will shift attention to the basal lamina and bottom of the tumour (AA1). As the animation will also lock attention, it was placed prior to the existing animation of the tumour spreading down into the connective tissue which otherwise might have competed for attention (AA3).

After the end of the new animation, the existing animation was then used to shift attention to the tumour spreading down into the connective tissue (AA1). With attention shifted to the spreading tumour, this animation was then paused half way through and a label revealed describing the animation. By pausing the animation, it should allow attention to move to the revealed label (AA3). After leaving time for the label to be read (TA2), the end of the animation was completed.

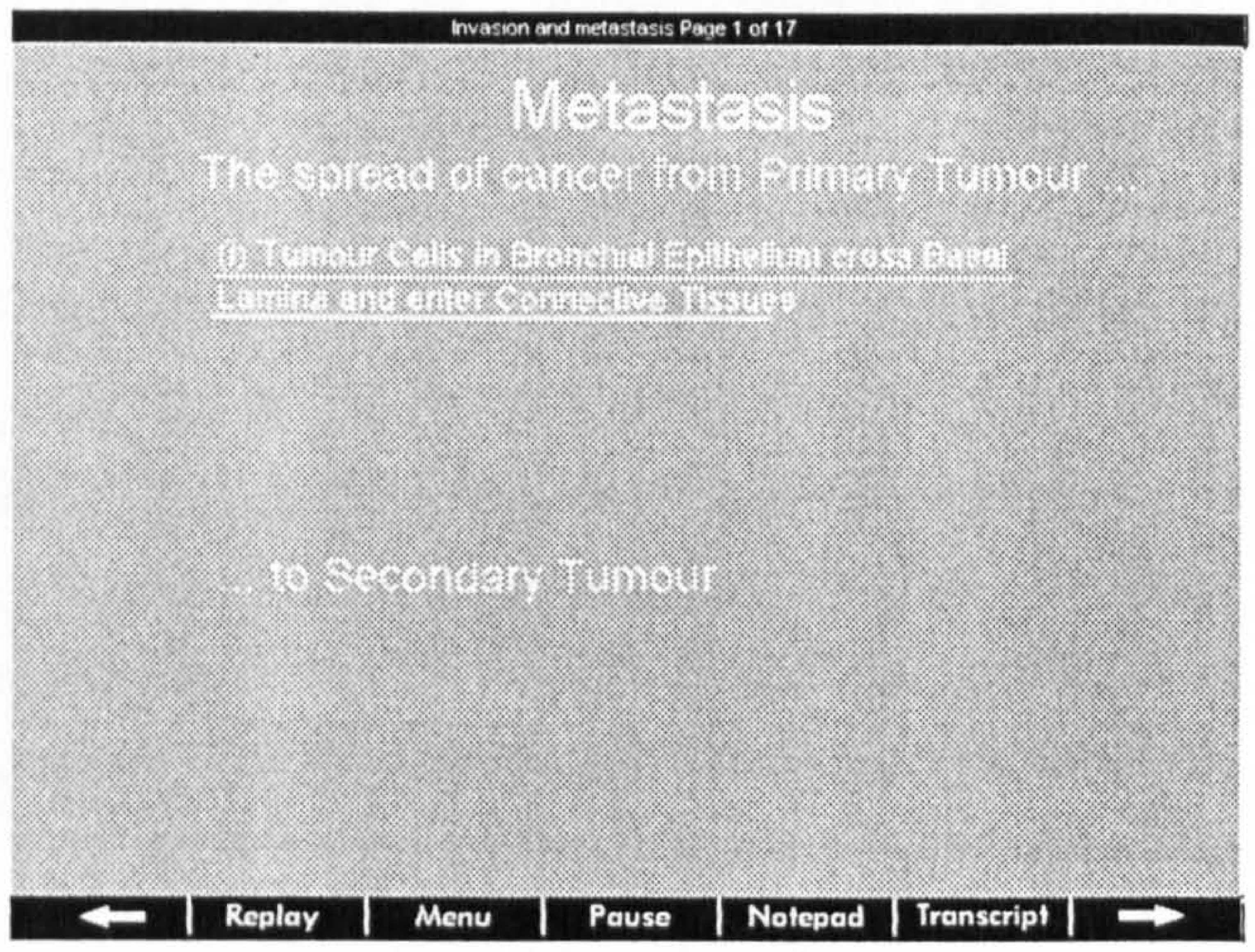
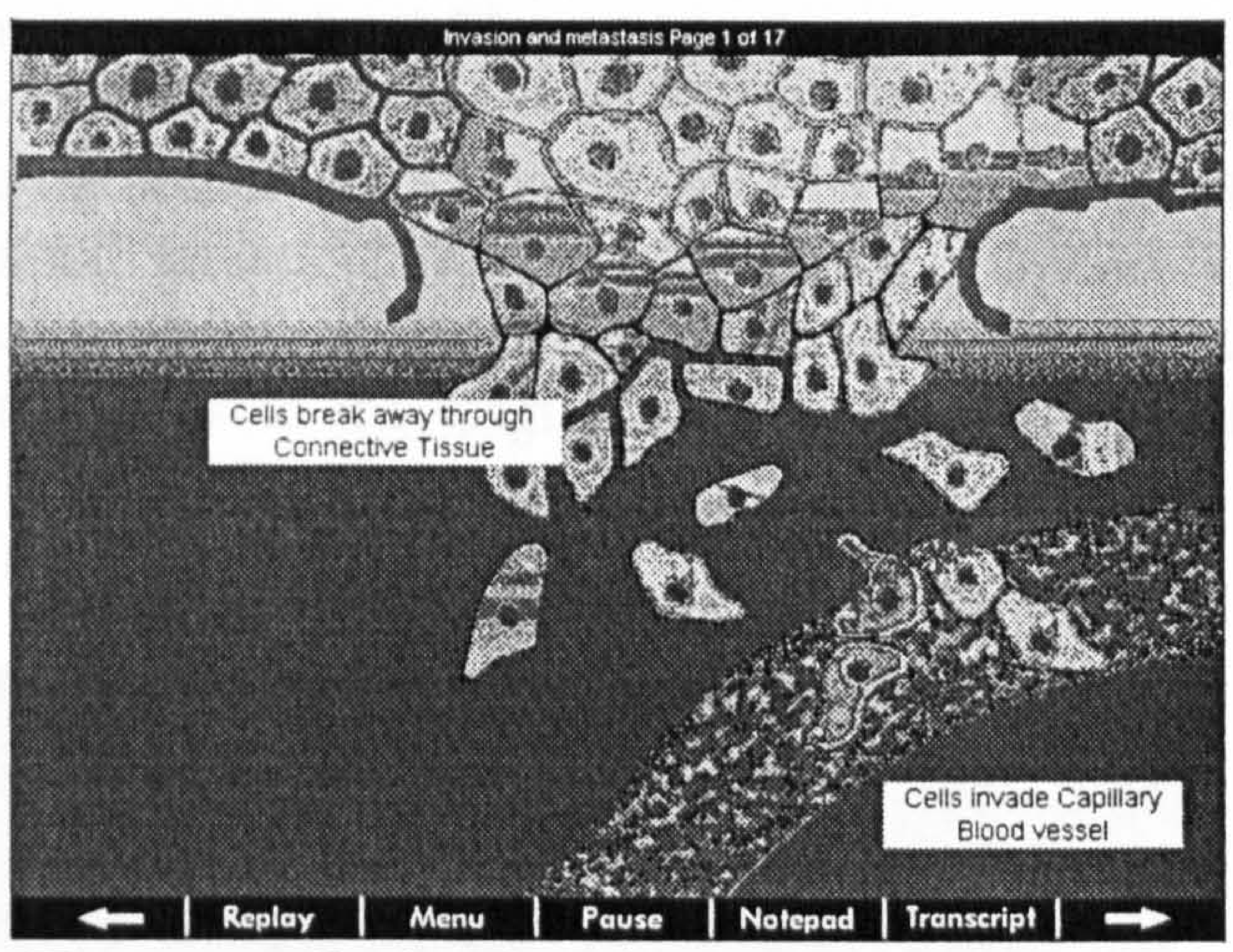


Figure 5.3.9

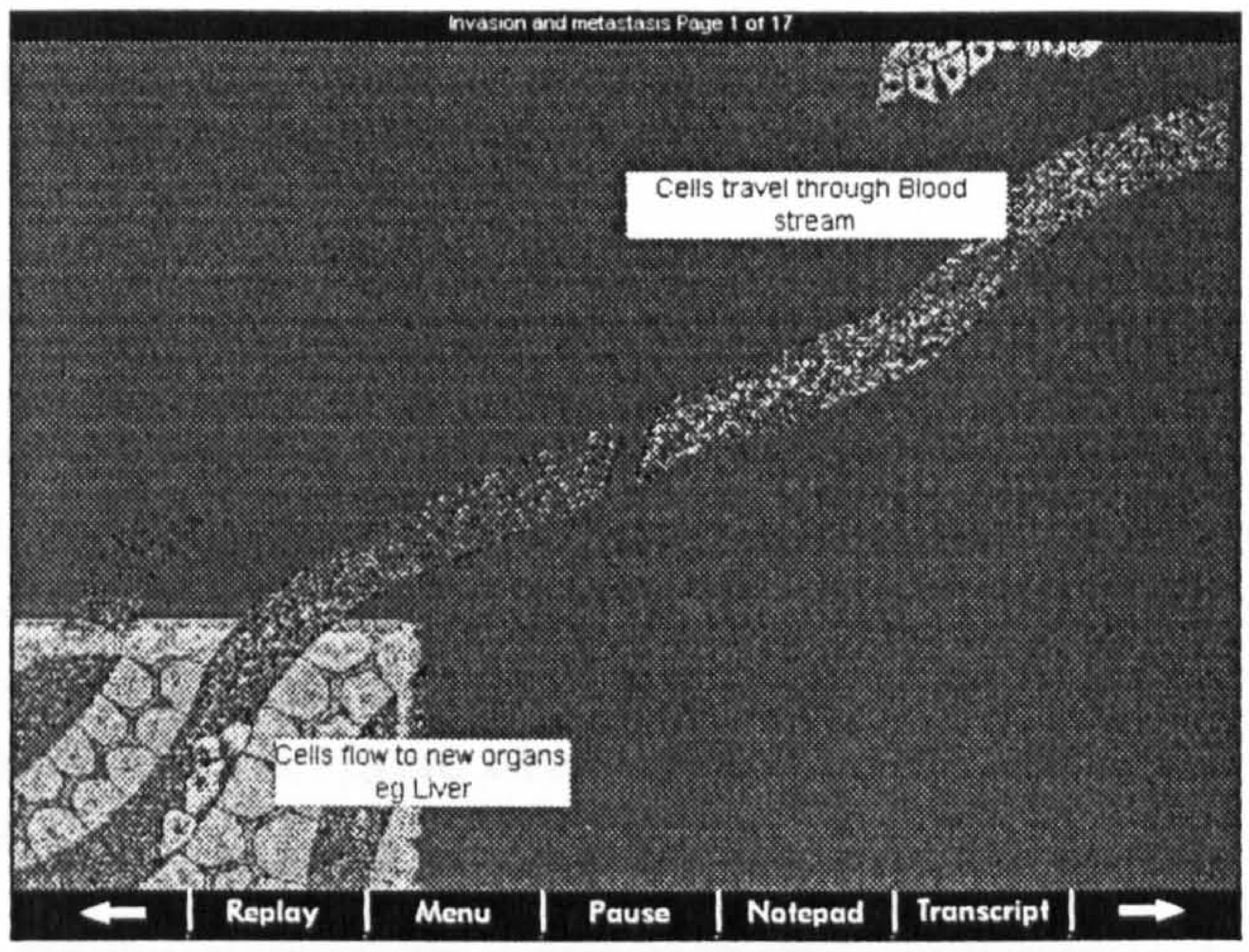
The next frame returned to the advanced organiser (LP2) and gave a summary to ensure the captions and animations were correctly integrated with the organiser (fig 5.3.9).



Tumour cells then break away through the connective tissues and invade a capillary blood vessel.

Figure 5.3.10

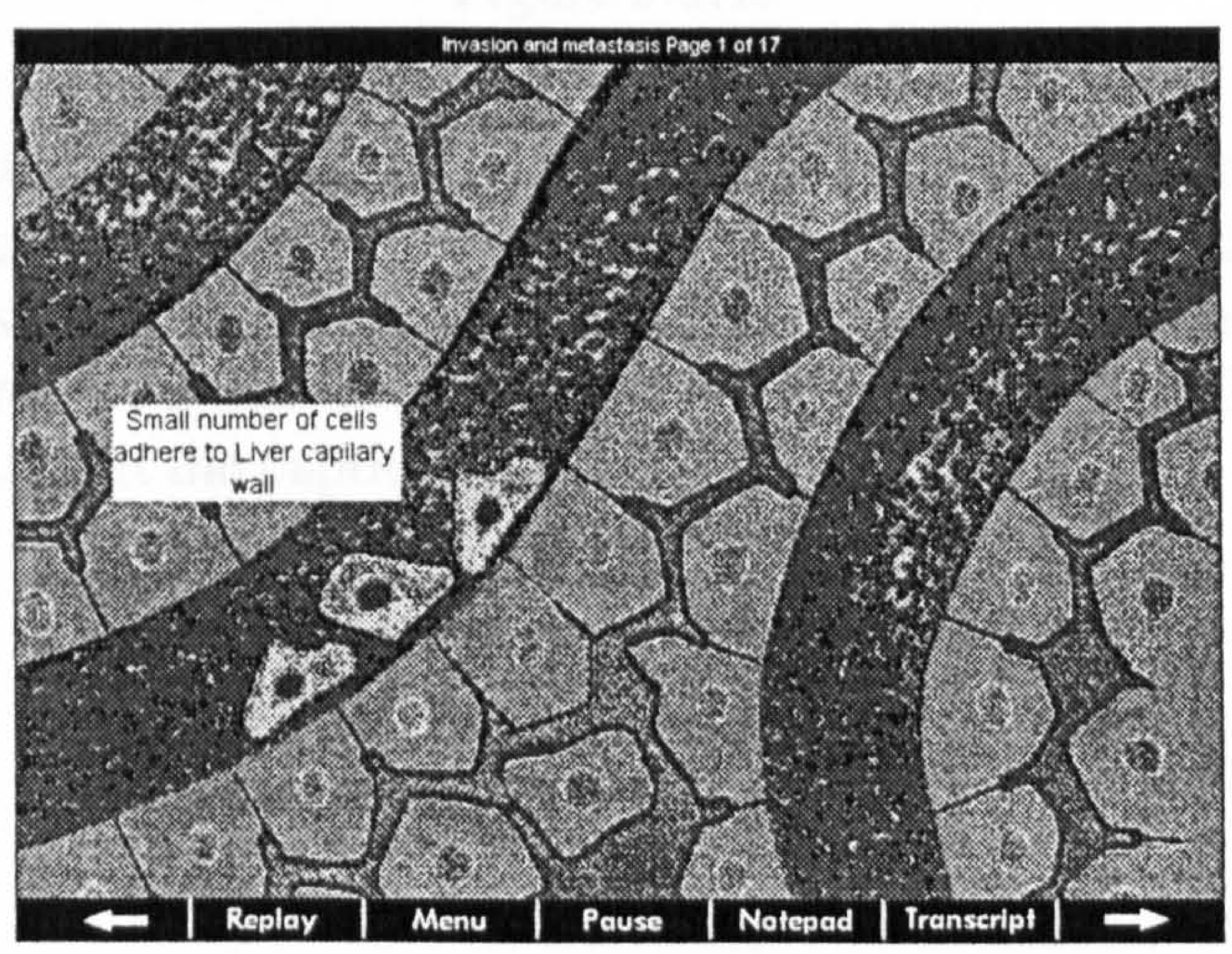
Following the organiser, the next frame used the original animation to show the tumour cells in motion toward the blood vessel (fig 5.3.10). Again, the metastasis label was removed to avoid diverting attention (IA7). The original animation was delayed to allow time to read the label 'Cells break through..' (AA2), (TA2). The animation was then shown. When it ended, the label 'Cells invade..' was revealed (IA5). This was left in its original position, since the animation would direct attention toward it. The speech track during the sequence was again modified to be in keeping with the labels shown (VR9).



The tumour cells now travel through the blood stream, away from the primary tumour.The cells flow to a new organ, the liver.

Figure 5.3.11

The organiser was then shown again (LP2), with a summary added of the previous frame 'Cells break through the capillary'. The following frame began with the original animation of the tumour cells moving through the blood vessel delayed to allow reading time for the label 'Cells travel through blood stream' (TA2), (AA2); again the metastasis label was also removed to avoid distracting attention (fig 5.3.11) (IA7). The animation was then shown, with the label revealed at its end, as in the original presentation. Again, the speech track was changed to match the labels (VR9); a temporal cue was also added 'now' to indicate the start of a new sequence (LP1).

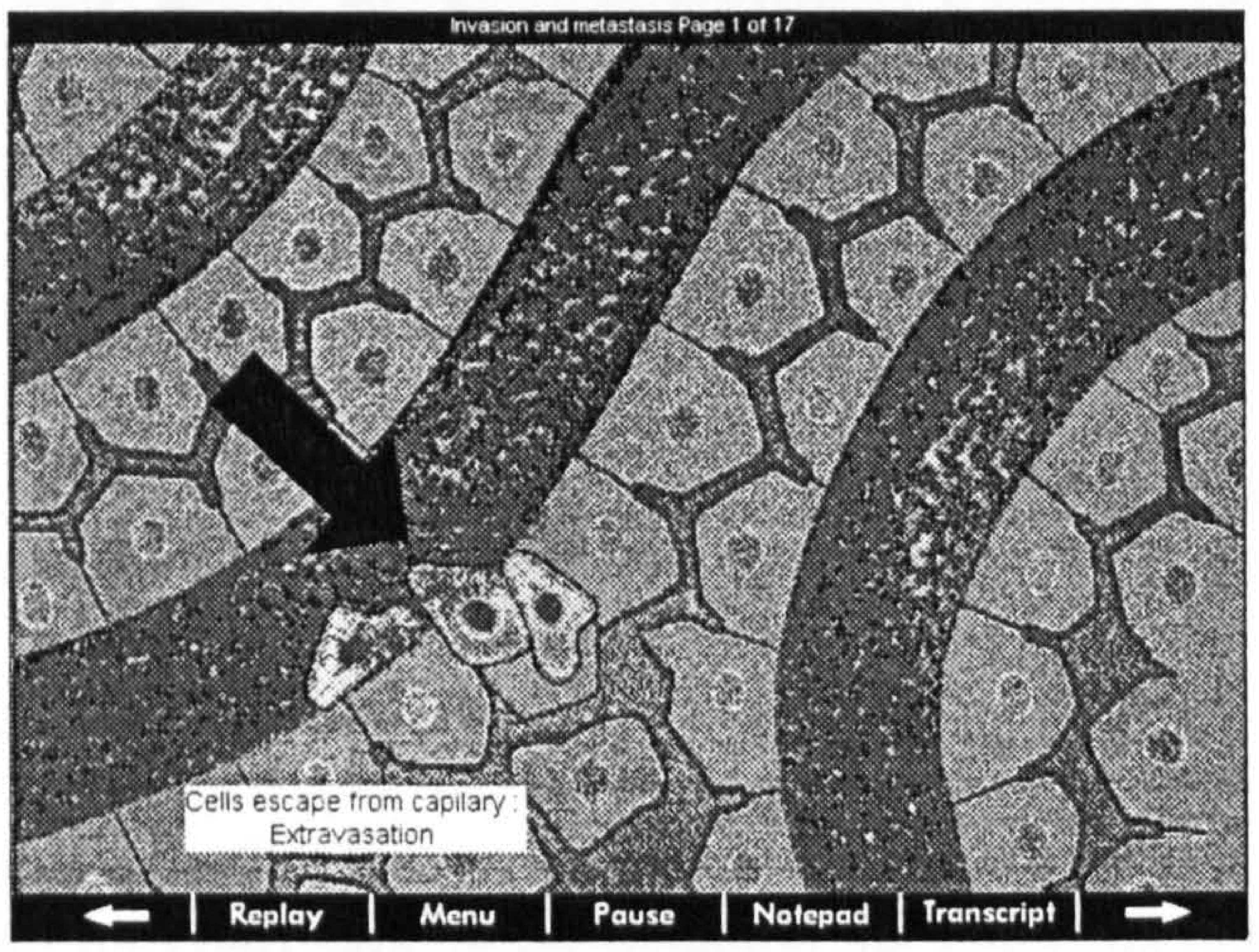


As they move through the liver, a small number of tumour cells may manage to adhere to the livers capillary wall.

Figure 5.3.12

The organiser was next shown and updated again to maintain the plan for the task. A zoom effect was then added to take the view from the tumour cells in the liver to a close up (AP4), focusing attention on the tumour cells, and providing context (AP6) as to where the tumour cells were positioned from the previous animation in the liver (fig 5.3.12).

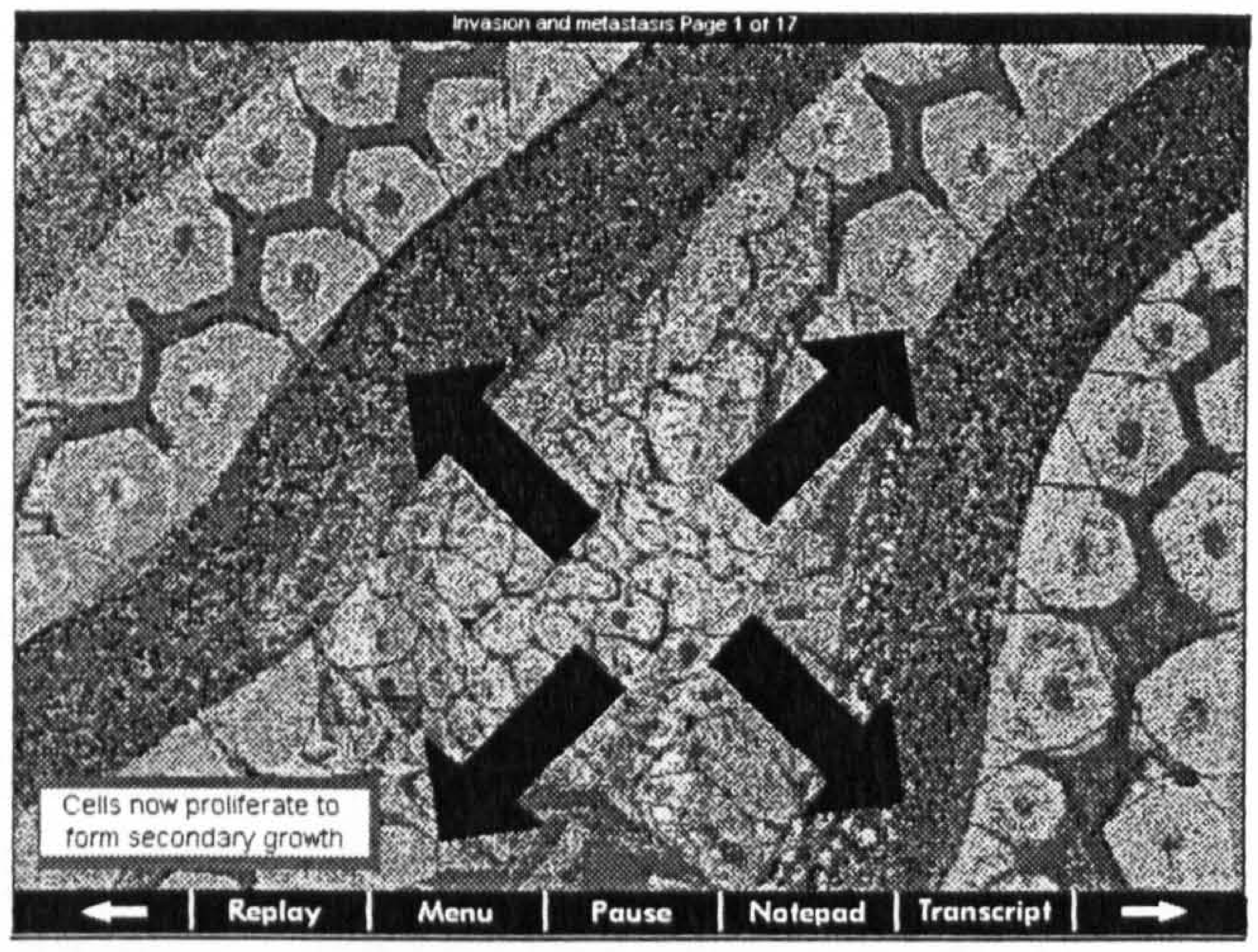
A new label and animation sequence were then added to show the process of the tumour cells attaching to the new site (SR5). The animation showed the tumour cells adhering to the cell wall. The speech was updated to support the 'attach' procedure (VR4). Another new animation was then added to support the 'extravisation' procedure, showing the process of the cells pushing into the surrounding tissues (SR5). An arrow was used to emphasise the tumour cells path into the tissues (IP8); with the original label adding grouping the actions of the tumour cells shown together as a single event (fig 5.3.13). The speech track was changed to support the label (VR9).



These tumour cells will then force an escape from the capillary, termed extravasation

Figure 5.3.13

The organiser was now again shown and further updated (LP2). The original animation of the tumour expanding was then shown, delayed to allow reading time for the caption (AA2), which was taken as-is from the original presentation. Arrows were added to emphasise the path of motion outwards (IP8), as the secondary growth occurs (fig 5.3.14). The speech track was also changed to support the caption (VR9).



Finally, the tumour cells proliferate to produce a secondary growth.

Figure 5.3.14

At the end of the sequence the organiser was shown (LP2), to provide an overview of the steps in the sequence and to organise them as a whole to support the 'metastasis' plan.

5.3.5 Re-authoring speech only condition

The speech & text only condition was based on the speech track from the re-authored presentation. It was split into two paragraphs breaking the sequence between the spread of the primary tumour and the formation of the secondary tumour.

Metastasis is the spread of cancer from primary to secondary tumours. The malignant primary tumour shown is in the patients bronchial epithelium. In order to spread, the cells of the primary tumour must break through the basal lamina and then enter the surrounding connective tissues. Tumour cells then break away through the connective tissues and invade a capilarly blood vessel.

The tumour cells now travel through the blood stream, away from the primary tumour. The cells flow to a new organ, the liver. As they move through the liver, a small number of tumour cells may manage to adhere to the livers capilarly wall. These tumour cells will then force an escape from the capilarly, termed extravasation. Finally, the tumour cells proliferate to produce a secondary growth.

Figure 5.3.15

5.3.6 Results for reauthored and speech only 'Metastasis' sequence Condition 3 & 4

Proposition type	Proposition	Condition 3 Re-authored Multimedia (/8)	Condition 4 Speech & Text (/8)
Cause	i) Metastasis is the spread of cancer.	5	6
Sequence	ii) Tumour cells break away from primary tumour.	7	3
Object state	iii) Malignant tumour cells are in bronchial epithelium	6	3
Action	iv) Tumour cells breakthrough basal lamina	5	1
Action	v) Tumour cells break away through the connective tissue	4	1
Sequence	vi) Tumour cells travel through the blood stream.	6	3
Action	vii) Tumour cells invade the blood vessel capillary	8	5
Action	viii) Tumour cells adhere to the capillary wall in liver	6	4
Action	ix) Tumour cells escape from capillary	5	4
Action	x) Tumour cells invade another organ.	6	1
Sequence	xi) A secondary tumour is formed.	7	6
	Total :	65	37

Table 5.3.3

5.3.7 Discussion for Condition 3 & 4

The result are shown in table 5.3.3. A significant difference was found between recall in the low domain knowledge group in the original design and the improved design (Mann-Whitney U, $P<0.001$). There was also a significant difference between recall of the high domain group and the improved design (Mann-Whitney U, $P<0.05$). The following improvements are of particular note :

- Recall of tumour cells being in the bronchial epithelium was greatly improved iii), over that given by the high knowledge subjects in the original presentation. This was re-authored using an improved caption and adding a reference to the object in the speech track.

- Recall of the tumour cells breaking through the basal lamina was greatly improved iv), over that given by the high knowledge subjects in the original presentation. This was re-authored as an animation, with an improved caption and highlight upon the basal lamina object. The speech track was also improved to reference the action.

- Recall of the tumour cells breaking through the connective tissue was improved vii). This was re-authored as an improved caption and a reference inserted into to the speech track.

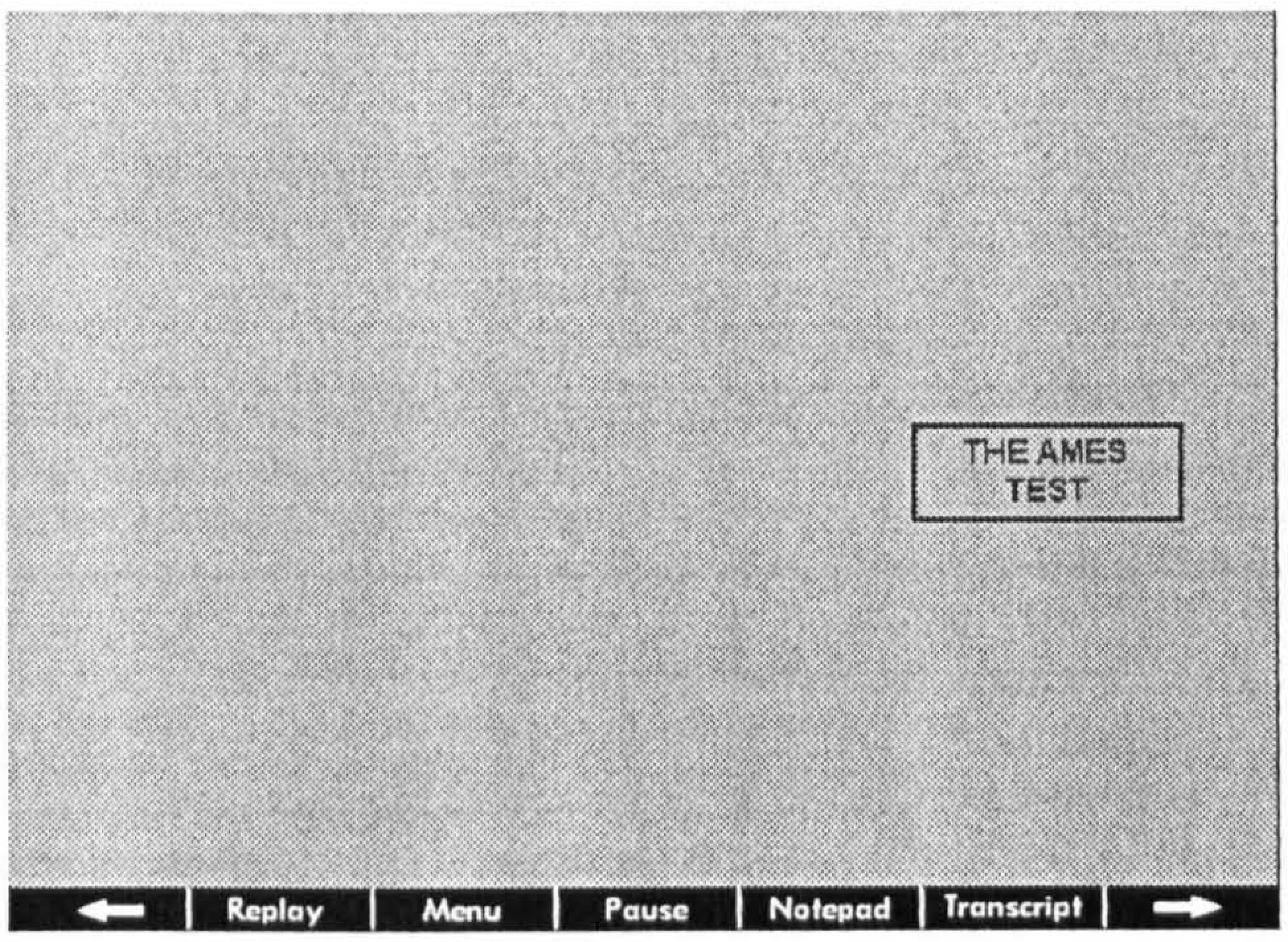
- Recall that the tumour cells invade the blood vessel capillary was improved viii). This was re-authored as an improved caption and a reference added into the speech track.

- Recall that the tumour cells adhere to the capillary wall was greatly improved ix). This was re-authored as a zoom-in effect to show the capillary wall in more detail, with an animation and symbol added to show the action of the tumour cell. An improved caption was also added, with a reference inserted into the speech track.

Recall performance with text-speech was significantly worse than for the low knowledge group in the improved multimedia design (Mann Whitney U, $P<0.01$), but was not significantly different from the low knowledge group for the original presentation. This showed that for re-authoring to provide additional linguistic information was not in itself sufficient to improve comprehension : animation and images were also vital for low knowledge subjects to understand the presentation.

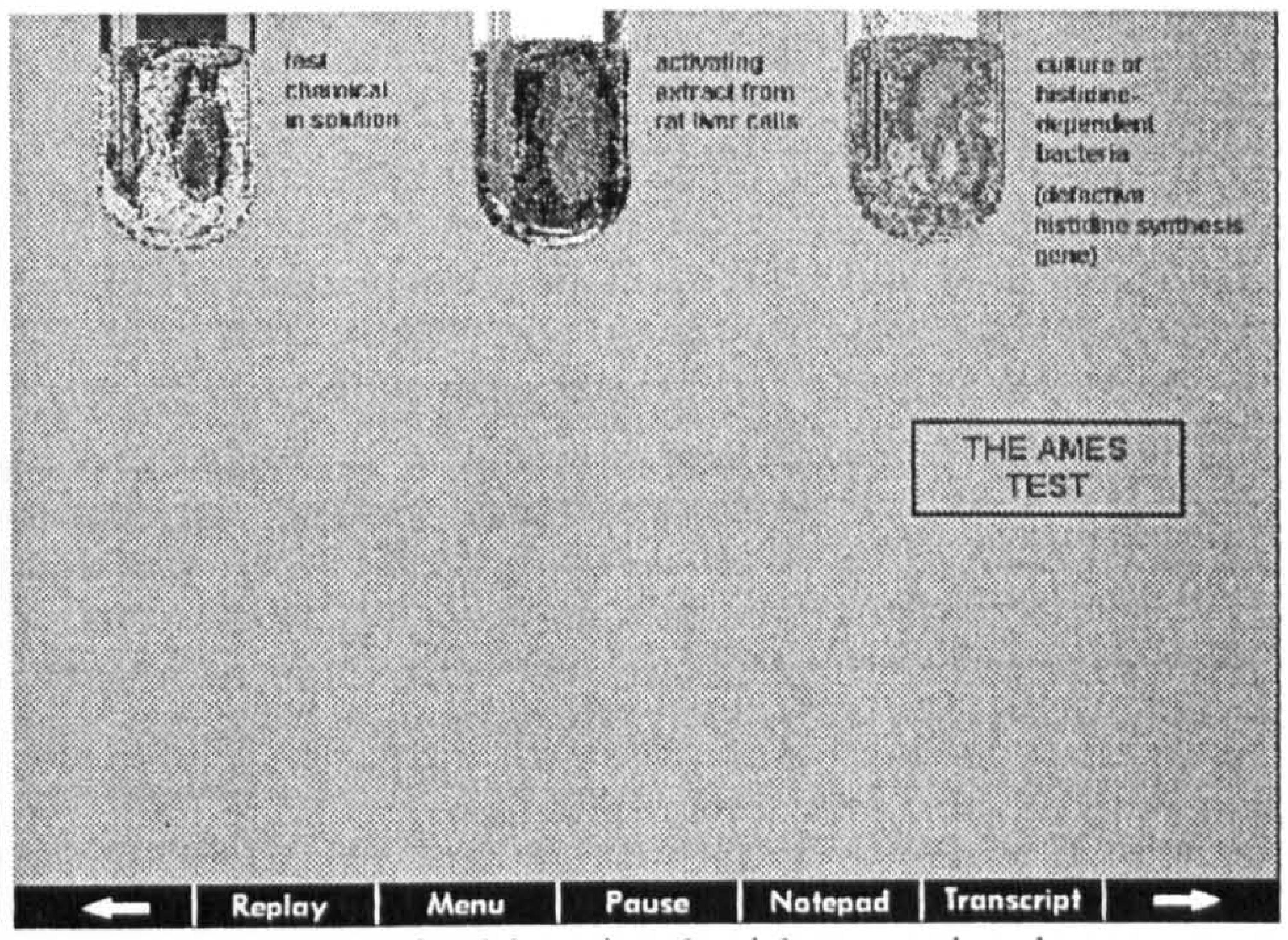
5.4 'Ames test' Sequence

5.4.1 Materials for 'Ames test' Sequence Condition 1 & 2



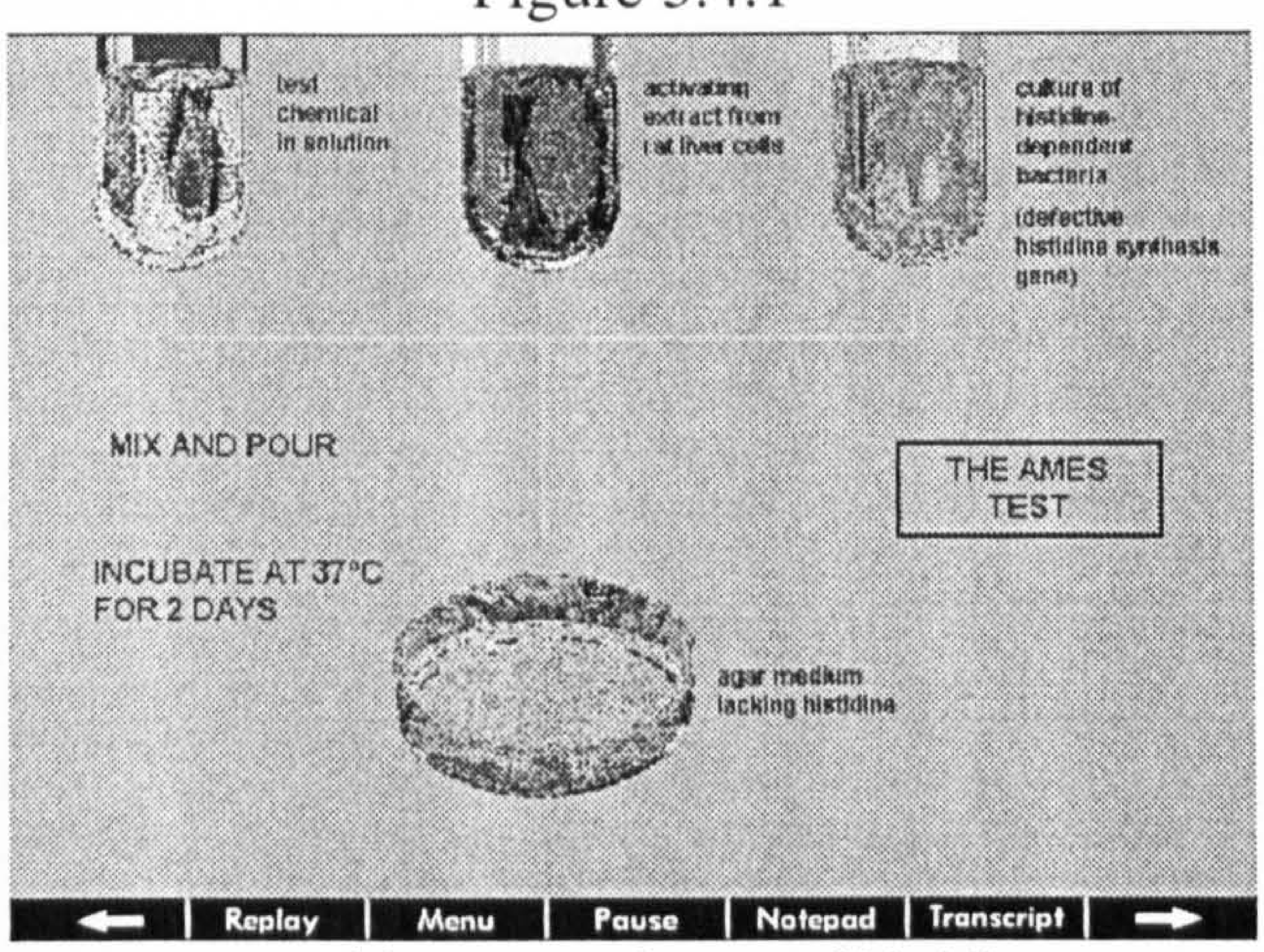
The most popular of these tests is the Ames test

Figure 5.4.1



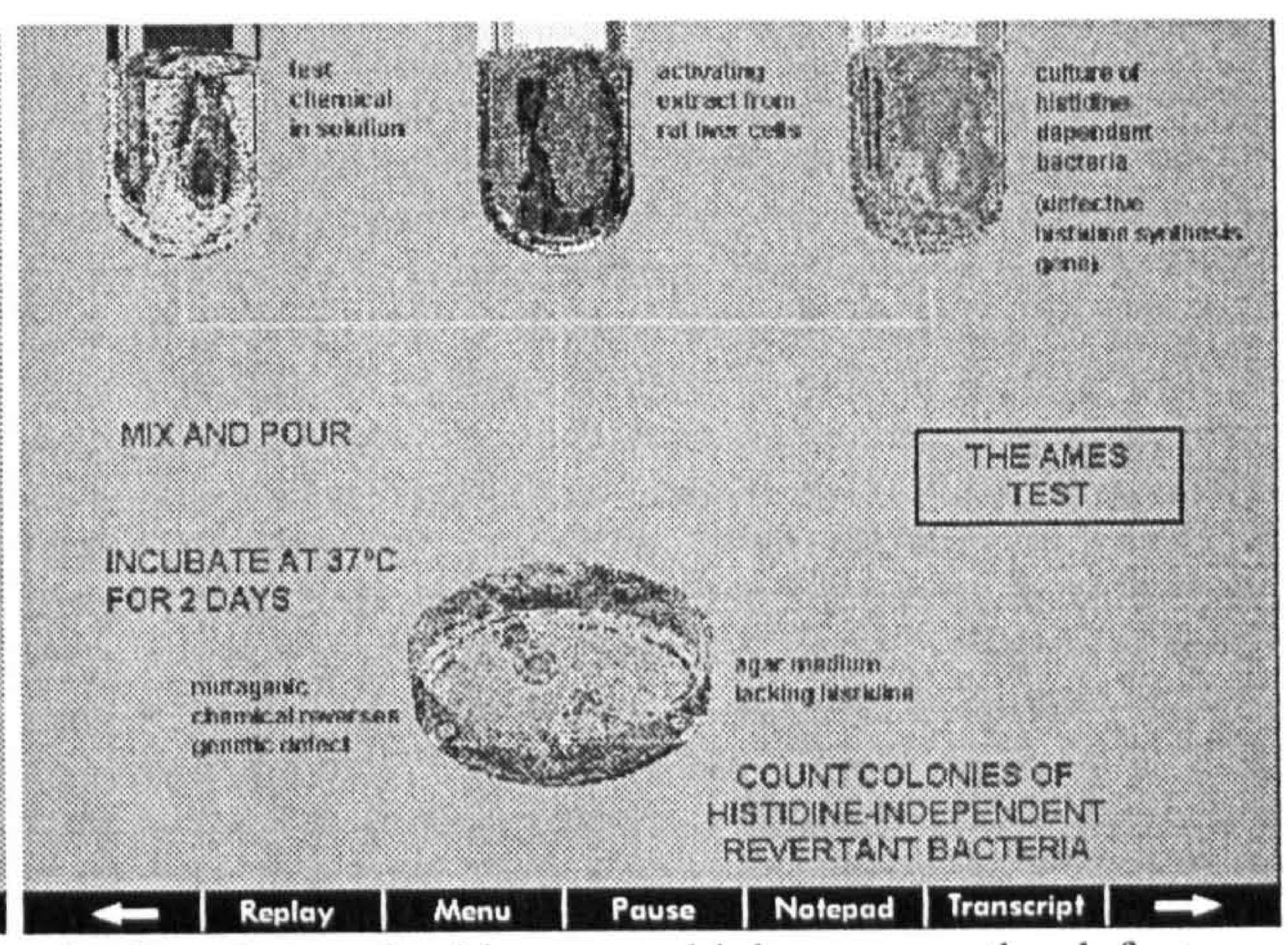
in which a chemical is mixed with an activating extract prepared from rat liver cells and added to a culture of specially designed bacteria which

Figure 5.4.2



cannot grow without a supplement of histidine, because they have a defect in a gene required for histidine synthesis. A mutagenic chemical can cause a

Figure 5.4.3



further change in this gene which reverses the defect, creating revertant bacteria which will grow without a histidine supplement and which can be counted as a measure of the mutagenic and therefore carcinogenic potential of the chemical under test.

Figure 5.4.4

The ames test presentation was divided into a set of ten propositions (figure 5.4.5). At the top level is a causal explanation, that mutagenic potential of a chemical can be used to measure carcinogenic potential. This is followed by three main parts, each given as a sequence. First a mixture is made using the histidine dependant bacteria. This requires that a set of chemicals are mixed together, then poured onto an agar plate and incubated, and that the agar plate has the state of having no histidine.

Next, if the test chemical is mutagenic it will reverse the defect in the histidine. This has two parts. First histidine independent bacteria are created, and then the bacteria grow on the agar plate. Finally the bacteria can be counted.

Problem

i) Measure the mutagenic and therefore carcinogenic potential of a chemical

- Sequence
  ii) Histidine dependant bacteria has a defective gene for histidine synthesis
  - Action
    iv) Mix the test chemical, rat liver cells and histidine dependant bacteria
  - Action
    v) Pour into an agar plat and incubate
  - Object state
    vi) Agar plate lacks histidinde
- Sequence
  vii) If the test chemical is mutagenic it will reverse the defect in the histidine dependant bacteria gene
  - Action
    viii)Hisitidine independent bacteria are created
  - Action
    viii)Hisitidine independent bacteria grow on the agar plate
- Sequence
  x)Histidine independent bacteria are counted

Figure 5.4.5

5.4.2 Results for 'Ames test' Sequence Condition 1 & 2

Proposition type	Proposition	Condition 1 Low Domain knowledge (/8)	Condition 2 High Domain knowledge (/8)	Total (/16)
Problem	i) Measure the mutagenetic and therefore carcinogenic potential of a chemical	7	5	12
Action	ii) Mix test chemical, activating extract of rat liver cells and histidine dependant bacteria	8	8	16
Sequence	iii) The histidine dependant bacteria has a defective gene for histidine synthesis	4	7	11
Action	iv) Pour onto an agar plate and incubate.	5	7	12
Object state	v) Agar plate lacks histidine.	1	3	4
Sequence	vi) If the test chemical is mutagenetic it will reverse the defect in the hisitidine dependant bacteria gene.	3	6	9
Action	vii) Histidine independent revertant bacteria are created.	0	2	2
Action	viii) The histidine independent bacteria will grow on the agar medium	3	7	10
Sequence	ix) The histidine independent bacteria can be counted.	5	5	10
	Total :	36	50	

Table 5.4.2

5.4.3 Discussion for Ames Test Condition 1 & 2

The results are shown in table 5.4.2. There was no significant difference in recall between high and low domain knowledge subjects, however the total recall scores for high domain were better than for low domain. This suggests that their performance was based only partly on what was given in the presentation, with background knowledge also aiding comprehension. This may have been the case because the sequence focused more on knowledge of chemical reactions than cancer.

- Recall of higher level sequence propositions i), vi), ix), representing a group of actions, was generally good for both high and low knowledge subjects.

- Use of captions and labels was inadequate in key areas. The central event of histidine dependant bacteria being created in vi), vii) could have been captioned more effectively e.g. giving the role of the test chemical and histidine dependant bacteria in the reaction, and providing details of the state change of the histidine dependant to independent bacteria. The caption gave 'mutagenic chemical reverses genetic defect', which neither identified the role of the test chemical and histidine independent bacteria, nor does it describe the state change. The caption containing histidine independent bacteria was not shown until the very end of the presentation, instead of when the reaction was described.

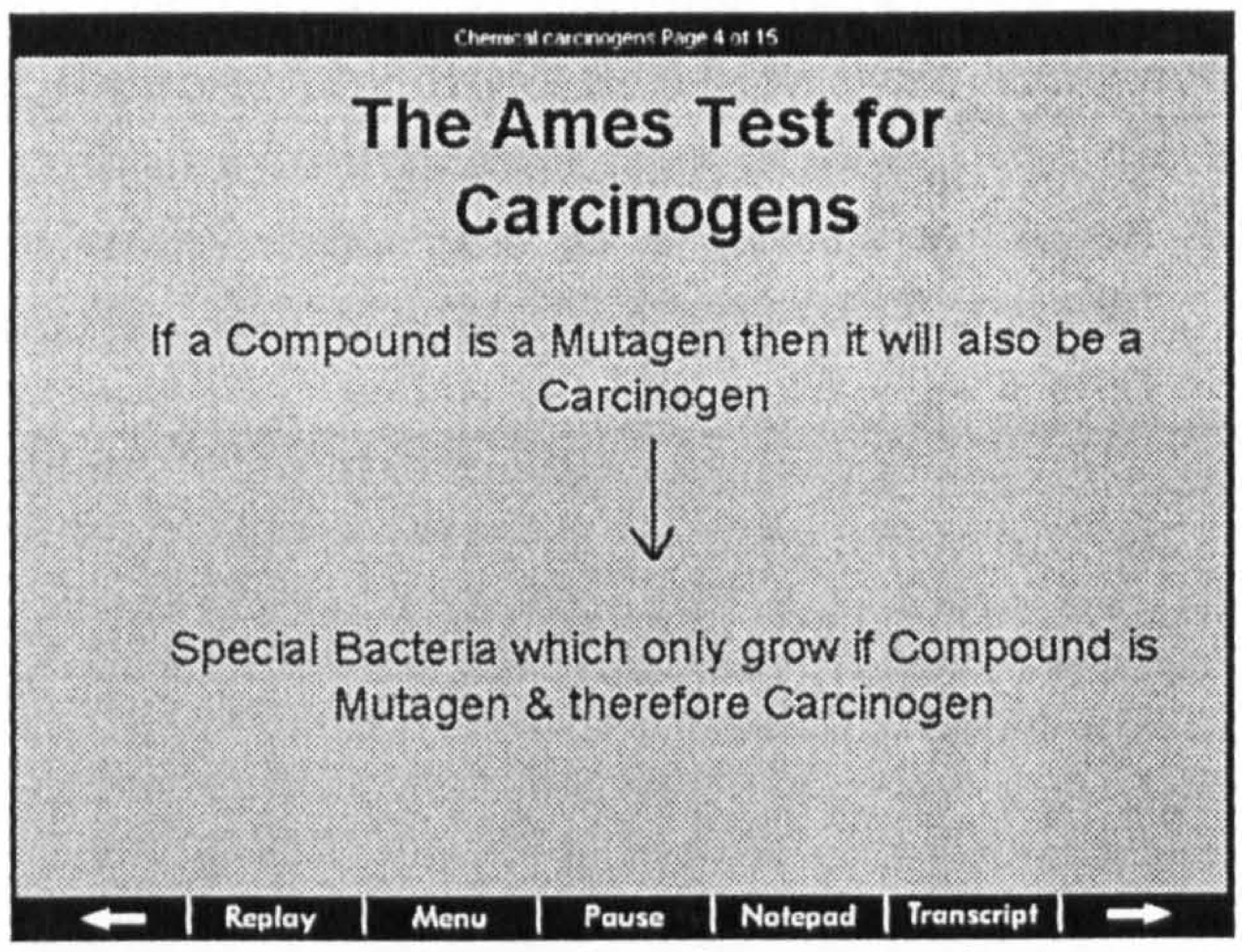
- Links between the speech track and objects may have been made overly complex by the large amount of text shown, and because wording between text and speech differed e.g. v) described 'agar medium lacked histidine' in text, whilst the speech gave 'added to a culture of specially designed bacteria which cannot grow without a supplement of histidine, because they have a defect in a gene required for histidine synthesis'. Similar differences in wording are also apparent in vi). Both propositions were poorly recalled.

- Whilst captions are very useful in supporting complex speech, it is also important to focus attention on particular captions when given in speech, and to use reveals to focus attention on particular captions. The histidine label in v) could have been revealed when cued by text. It may have been useful to refocus attention on the test chemical and hisitidine dependant bacteria during description of the reaction when they were cued, using an emphasise effect. The ames test label could also be removed to reduce the clutter.

- Animation could have been more effectively used to show the process of the test chemical changing the state of histidine dependant bacteria. This may have improved recall of vi),vii) particularly for the low domain knowledge subjects. Animation could have included showing the reaction of the two chemicals in the agar. Animation may be particularly useful in showing complex reactions. The animation used showed only the final growth of the independent bacteria.

- Where the speech track primed or re-inforced the propositions presented visually, comprehension was good. This was particularly the case for ii) describing the components of the test, iv) mix and pour, ix) count bacteria.

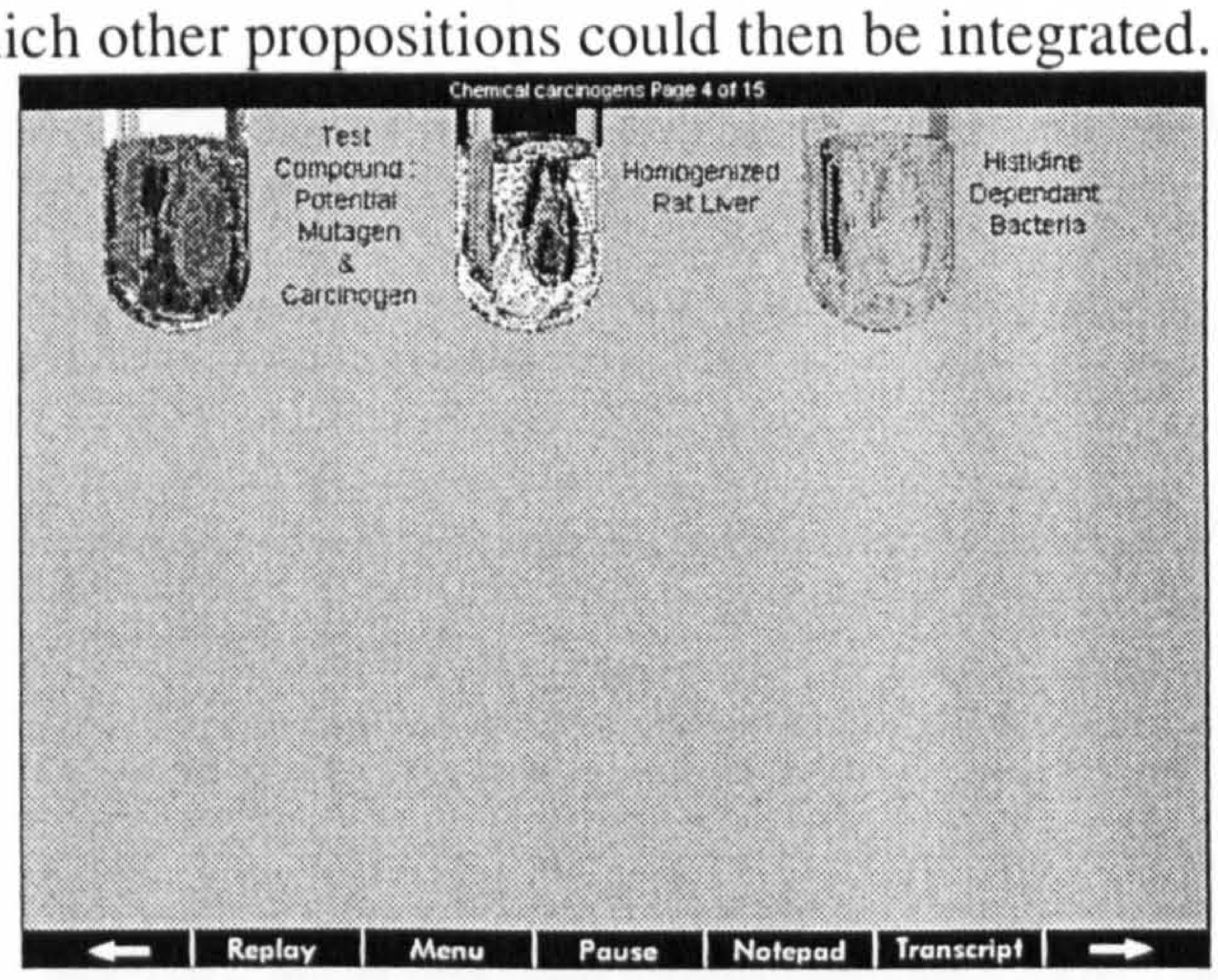
5.4.4 Re-authoring 'Ames test' Sequence for Condition 3



The ames test is used to detect carcinogens. It uses the fact that if a compound is a mutagen then it will also be a carcinogen.

Figure 5.4.6

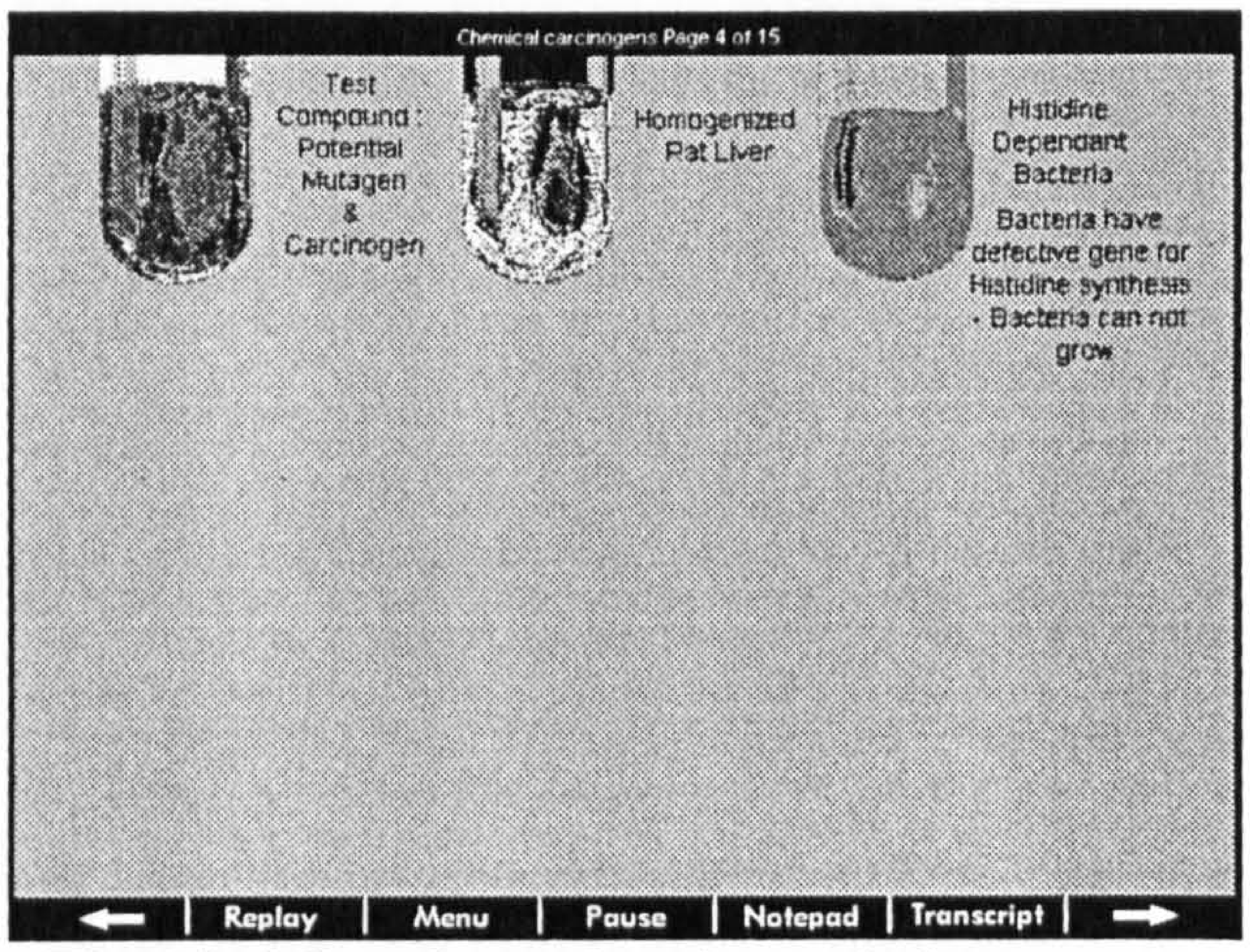
The redesign began by providing an advanced organiser for the cause-effect relation which provided the plan of the sequence (LP2), using a caption and speech track which introduced the main topic of the Ames test (fig 5.4.6). An arrow symbol was used to show the causal relation. The caption was tightly linked to the speech track to reinforce the propositions given (VR5). The organiser should form the top level of the mental model for this part of the presentation, below which other propositions could then be integrated.



The test requires a special bacteria which will only grow if a compound is a mutagen, and therefore carcinogen. The test compound which is the potential mutagen and therefore carcinogen, is mixed with an activating extract of rat liver cells, and with the specially designed histidine dependant bacteria.

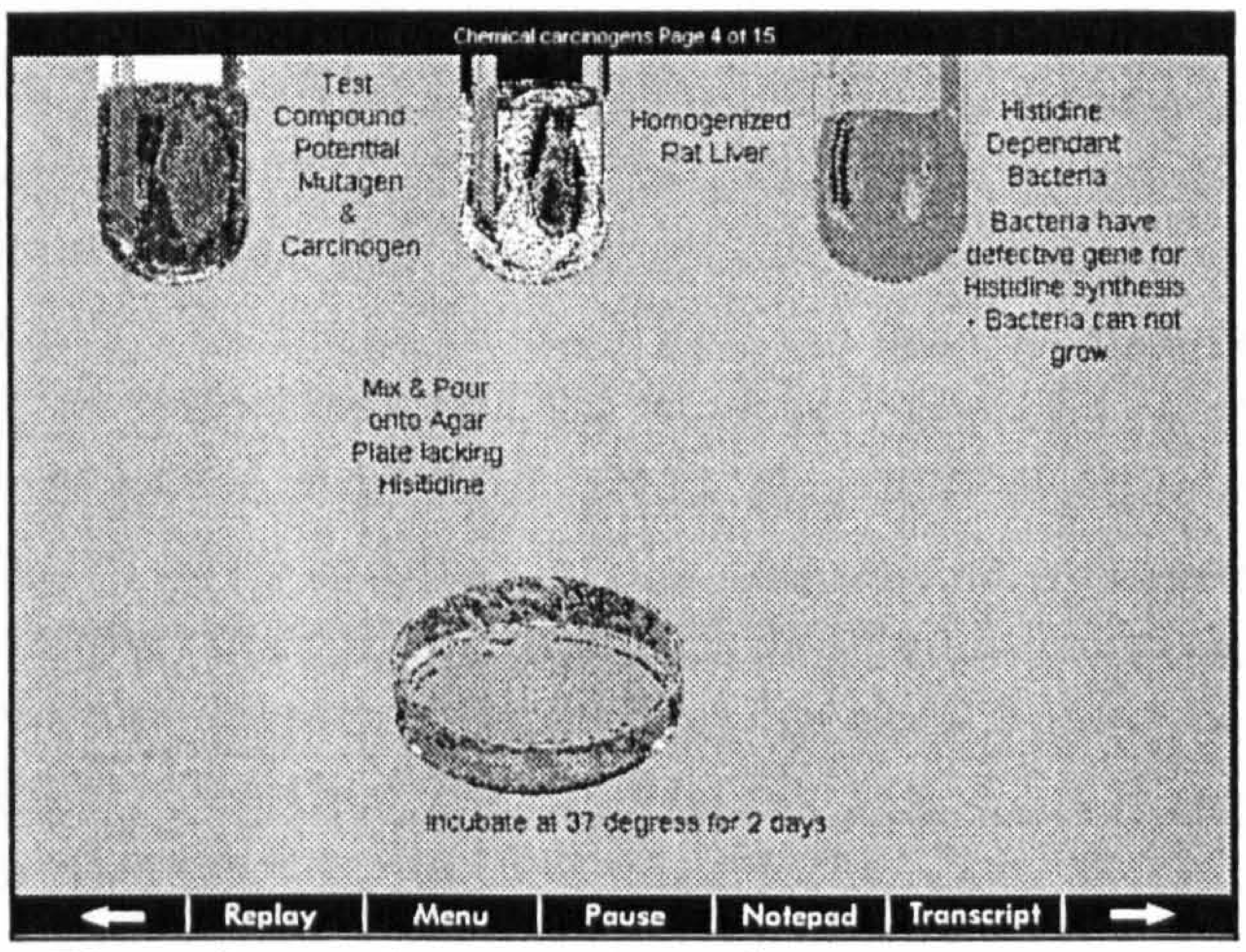
Figure 5.4.7

The opening sequence was taken almost verbatim from the original presentation (fig 5.4.7). The attributes of the test compound and histidine dependant bacteria were stressed in the speech track (SA4), and supported by captions which were more detailed than in the original presentation, since the complex information may be lost from the speech track (VR3). The test chemical was additionally captioned as 'Potential mutagen and carcinogen', and the hisitidine dependant bacteria caption added 'Bacteria can not grow'.



The histidine dependant bacteria have a defect in the gene for histidine synthesis. This gene is required for the bacteria to grow.

Figure 5.4.8

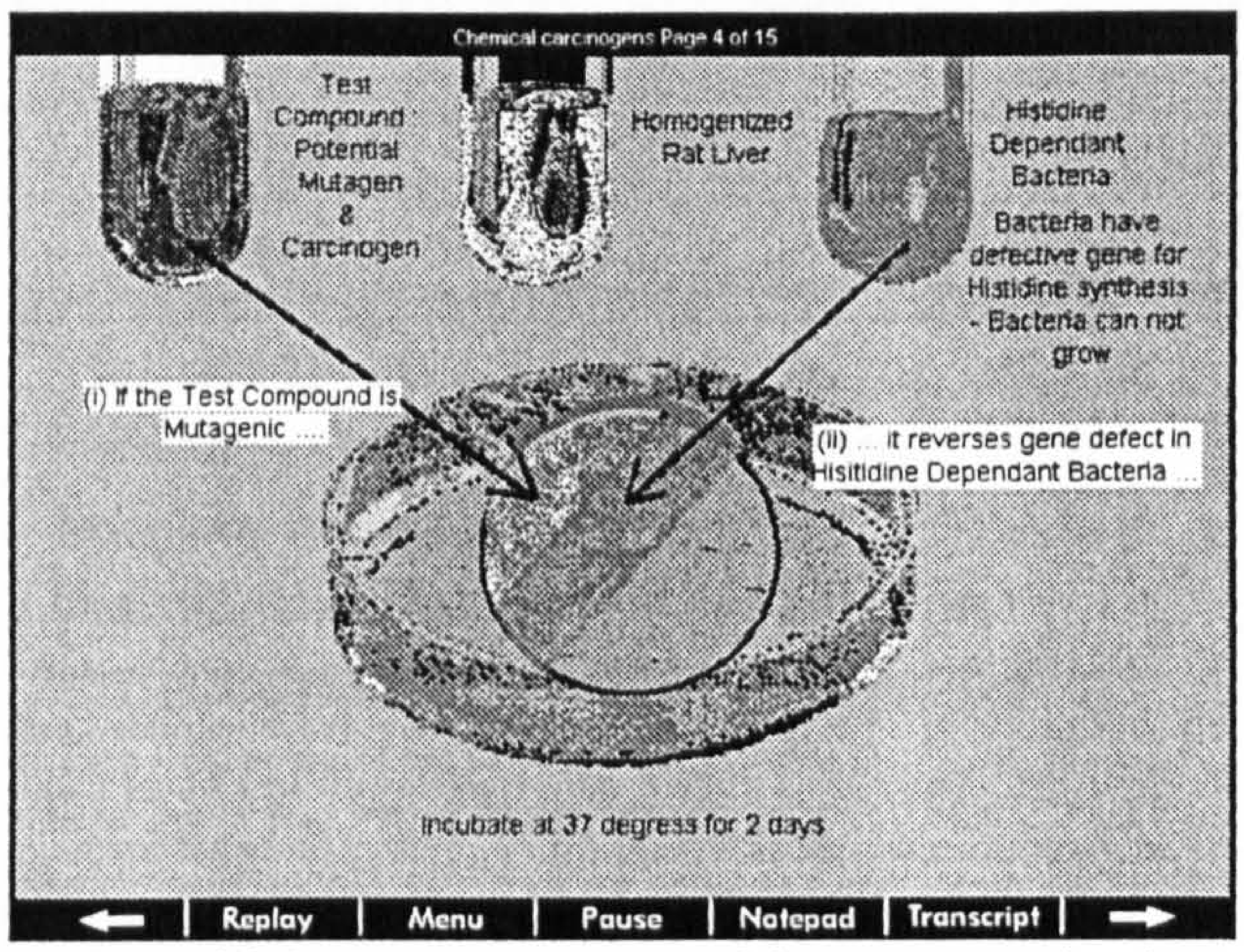


The mixture is now poured onto an agar plate. The agar plate lacks histidine, so the hisitidine dependant bacteria cannot grow.

Figure 5.4.9

An animation was then added to emphasise the state of the histidine dependant bacteria (SR5) (AA1), showing it change state when the speech cued that it has 'a defect for histidine synthesis'; the histidine dependant bacteria changed colour and shape to indicate the defect produced a state change (fig 5.4.8). Additional text was also shown to re-inforce the speech track for this proposition (VR3).

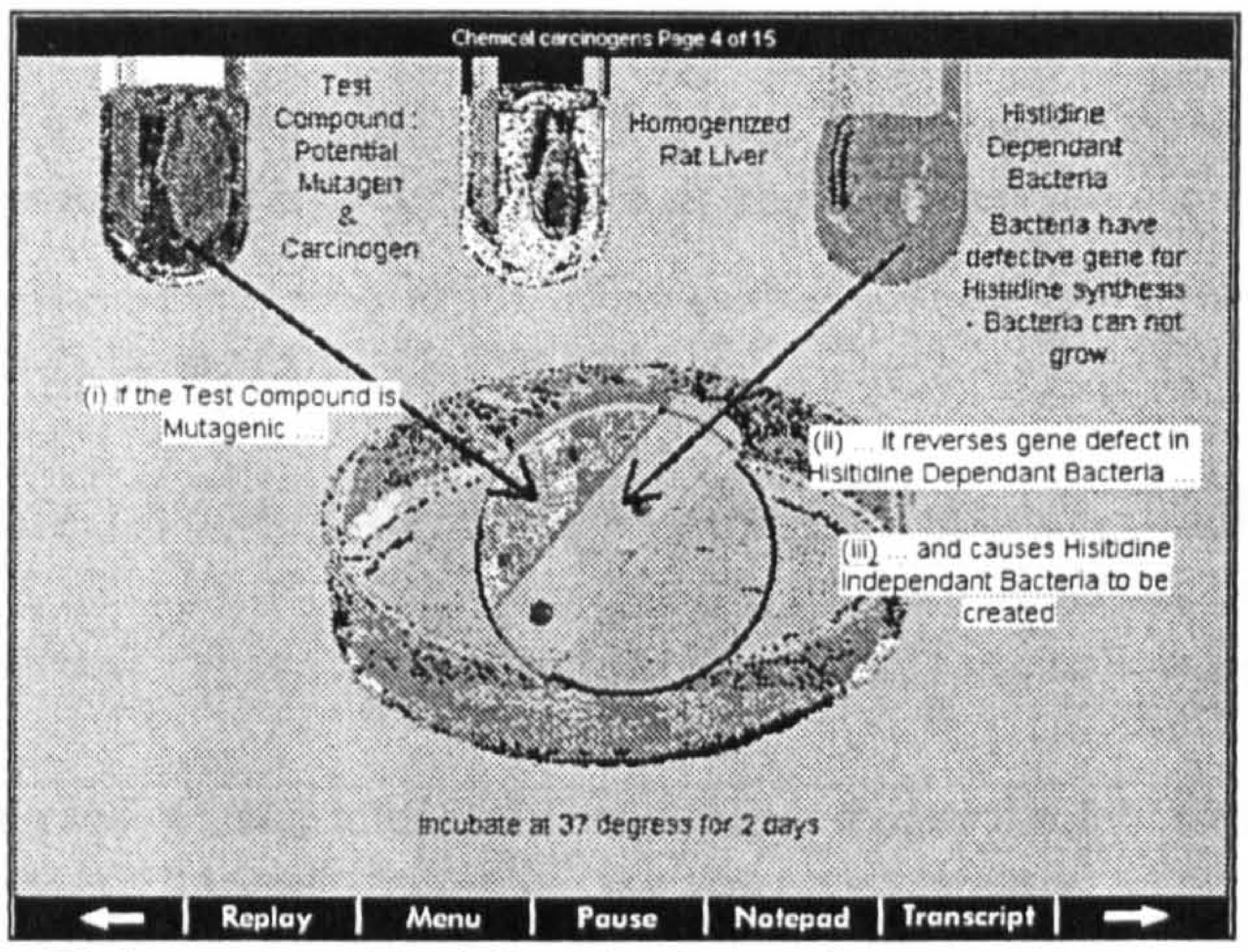
The original presentation was then used to show the agar plate, with the speech track changed to support the original label 'agar plate lacking in histidine' (VR9) (fig 5.4.9). The original 'incubate' label was then shown, and a pause allowed for reading time for the two labels (TA2).



Now, if the test compound is mutagenic, it will reverse the gene defect in the hisitidine dependant bacteria . This causes histidine independent revertant bacteria to be created.

Figure 5.4.10

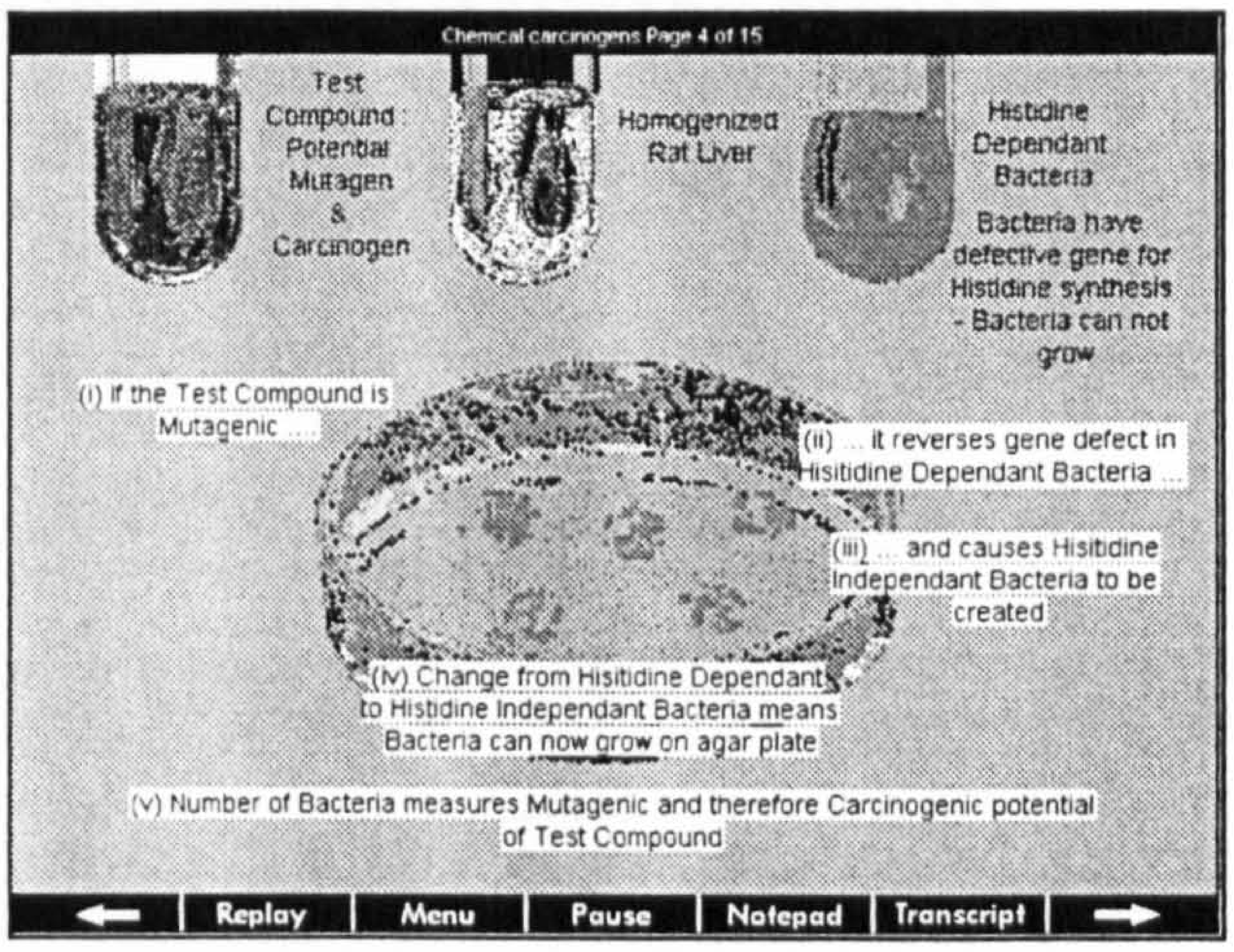
A zoom effect (AP5) was now used to focus attention on the agar plate (fig 5.4.10). A label was then revealed, moving attention to the test compound (IA5), which the speech cued 'Now, if the test compound is mutagenic..' (CP3). The speech provided a cue to the sequence order, 'Now..' (LP1). At this point, a line was used to link the test compound to the graphic of the agar plate (CP1), to shift attention between the plate and test chemical to support the speech cue to 'test compound'. A pause was then given to allow reading time (TA2). This was followed by another revealed label and line (CP1), linking the histidine dependant bacteria to the plate. This shifted attention over to the histidine dependant bacteria when the speech cued 'it will reverse gene defect in histidine dependant bacteria' (CP6). Reading time was then allowed for the label (TA2). Together, the two labels and speech explained the role of the objects in the action of producing histidine independent bacteria (SR6).



This causes histidine independent revertant bacteria to be created. This change from histidine dependant to histidine independent bacteria means that the bacteria can now grow on the agar plate.

Figure 5.4.11

An animation was then used to show a reaction on the agar plate (SR5), with the colour and appearance of the agar medium changing in places to show the state change to histidine independent bacteria (fig 5.4.11). The speech provided the object role in this state change and used a causal cue 'This causes histidine independent bacteria to be created' to provide a cause-effect relation to the animation (LP3). This was supported by a caption, numbered to support the order of the sequence (LP2).



The bacteria can now be counted. The number of bacteria measures the mutagenetic and therefore carcinogenic potential of the test compound.

Figure 5.4.12

The final frame showed the histidine dependant bacteria on the plate, as in the original presentation (fig 5.4.12). The original speech track was used, but a caption was added to support it because it provided an important part of the sequence.

5.4.5 Re-authoring speech & text only Condition 4

The speech & text only condition was based on the speech track from the re-authored presentation. It was split into four paragraphs giving an introduction and then representing the three main phases of the test : mix chemicals, incubate and grow, and count bacteria.

The ames test is used to detect carcinogens. It uses the fact that if a compound is a mutagen then it will also be a carcinogen.

The test requires a special bacteria which will only grow if a compound is a mutagen, and therefore carcinogen. The test compound which is the potential mutagen and therefore carcinogen, is mixed with an activating extract of rat liver cells, and with the specially designed histidine dependant bacteria. The histidine dependant bacteria have a defect in the gene for histidine synthesis. This gene is required for the bacteria to grow.

The mixture is now poured onto an agar plate. The agar plate lacks histidine, so the hisitidine dependant bacteria cannot grow. The agar plate and mixture are now incubated at 37 degrees for 2 days. Now, if the test compound is mutagenetic, it will reverse the gene defect in the hisitidine dependant bacteria . This causes histidine independant revertant bacteria to be created.

This change from histidine dependant to histidine independant bacteria means that the bacteria can now grow on the agar plate. The bacteria can now be counted. The number of bacteria measures the mutagenetic and therefore carcinogenic potential of the test compound.

Figure 5.4.13

5.4.6 Results for reauthored and speech only 'Ames test' Sequence Condition 3 & 4

Proposition type	Proposition	Condition 3 Re-authored Multimedia (/8)	Condition 4 Text & speech only (/8)
Problem	i) Measure the mutagenetic and therefore carcinogenic potential of a chemical	6	7
Action	ii) Mix test chemical, activating extract of rat liver cells and histidine dependant bacteria	7	4
Sequence	iii) The histidine dependant bacteria has a defective gene for histidine synthesis	6	1
Action	iv) Pour onto an agar plate and incubate.	7	7
Object state	v) Agar plate lacks histidine.	2	2
Sequence	vi) If the test chemical is mutagenetic it will reverse the defect in the hisitidine dependant bacteria gene.	7	0
Action	vii) Histidine independent revertant bacteria are created.	5	0
Action	viii) The histidine independent bacteria will grow on the agar medium	6	0
Sequence	ix) The histidine independent bacteria can be counted.	6	3
	Total :	52	24

Table 5.4.3

5.4.7 Discussion for 'Ames Test' Condition 3 & 4

The results are shown in table 5.4.3. Total recall scores were higher for the re-authored condition than for low or high domain subjects in the original presentation. However, the difference was non-significant difference between recall in the low domain knowledge group in the original design and the improved design (Mann-Whitney U, $P<0.09$). This failure may be due to the relatively good recall that the original Ames test sequence produced, as only propositions v), vi), vii), viii) were extremely poorly recalled in the original. Re-authoring improved their recall. The following improvements are of particular note :

- Recall of histidine dependant bacteria having a defective gene iii) was slightly improved. This was re-authored as an animation to show the gene defect, with additional detail given in speech and the text caption to support it.

- Recall of a mutagenic test chemical reversing the defect in the histidine dependant bacteria was greatly improved vi). This was re-authored using arrow symbols to show the role of the test chemical upon the histidine dependant bacteria. Additional details were also provided in the speech and text captions were added to re-inforce it's role.

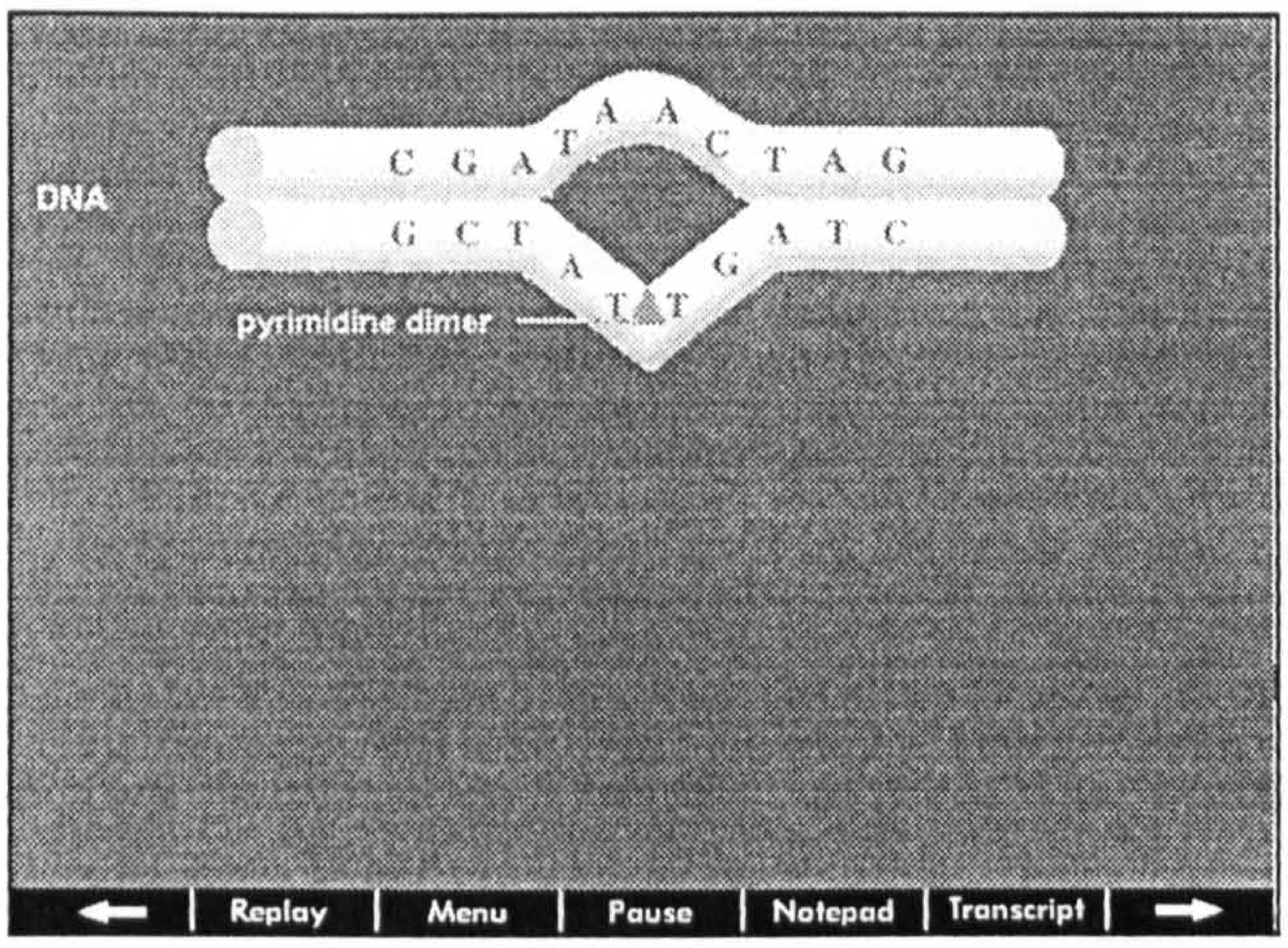
- Recall of histidine dependant bacteria being formed was greatly improved vii). This was shown as animation, changing the appearance of the histidine dependant bacteria due to the test chemical. This was re-inforced by the use of arrow symbols to relate the change from histidine dependant to independent bacteria and by the use of an additional caption and speech.

- Recall of histidine dependant bacteria growing on agar plate was improved viii). This was given as a caption and added to the speech track during the animation of the histidine independent bacteria growing.

Recall performance with text-speech was significantly worse than for the low knowledge group in the improved multimedia design (Mann Whitney U, $P<0.001$), but was not significantly different from the low knowledge group for the original presentation. This again showed that for re-authoring to provide additional linguistic information was not in itself sufficient to improve comprehension : animation and images were also vital for low knowledge subjects to understand the presentation.

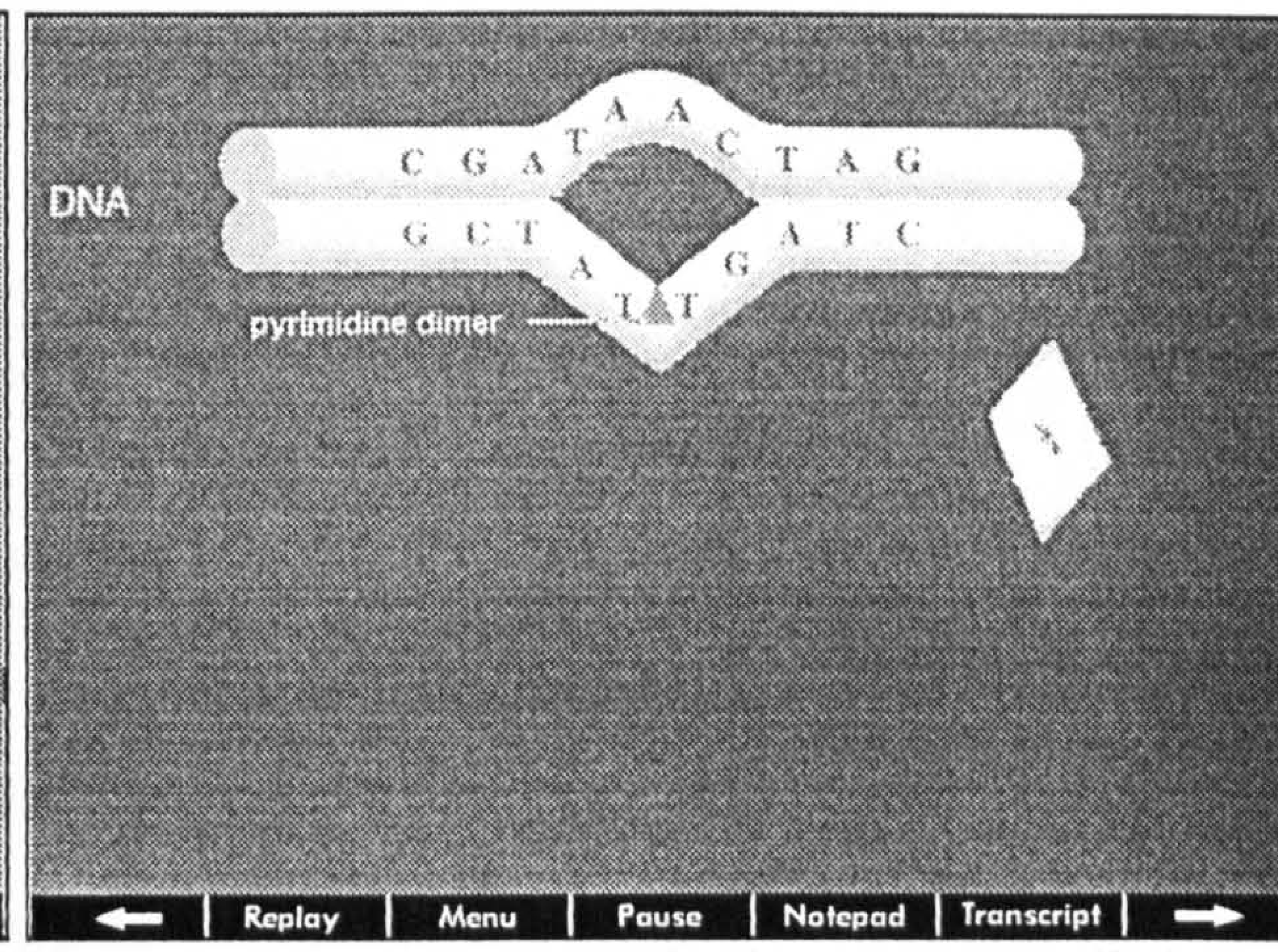
5.5 'Photoreactivation' sequence

5.5.1 Materials for 'Photoreactivation' sequence Condition 1 & 2



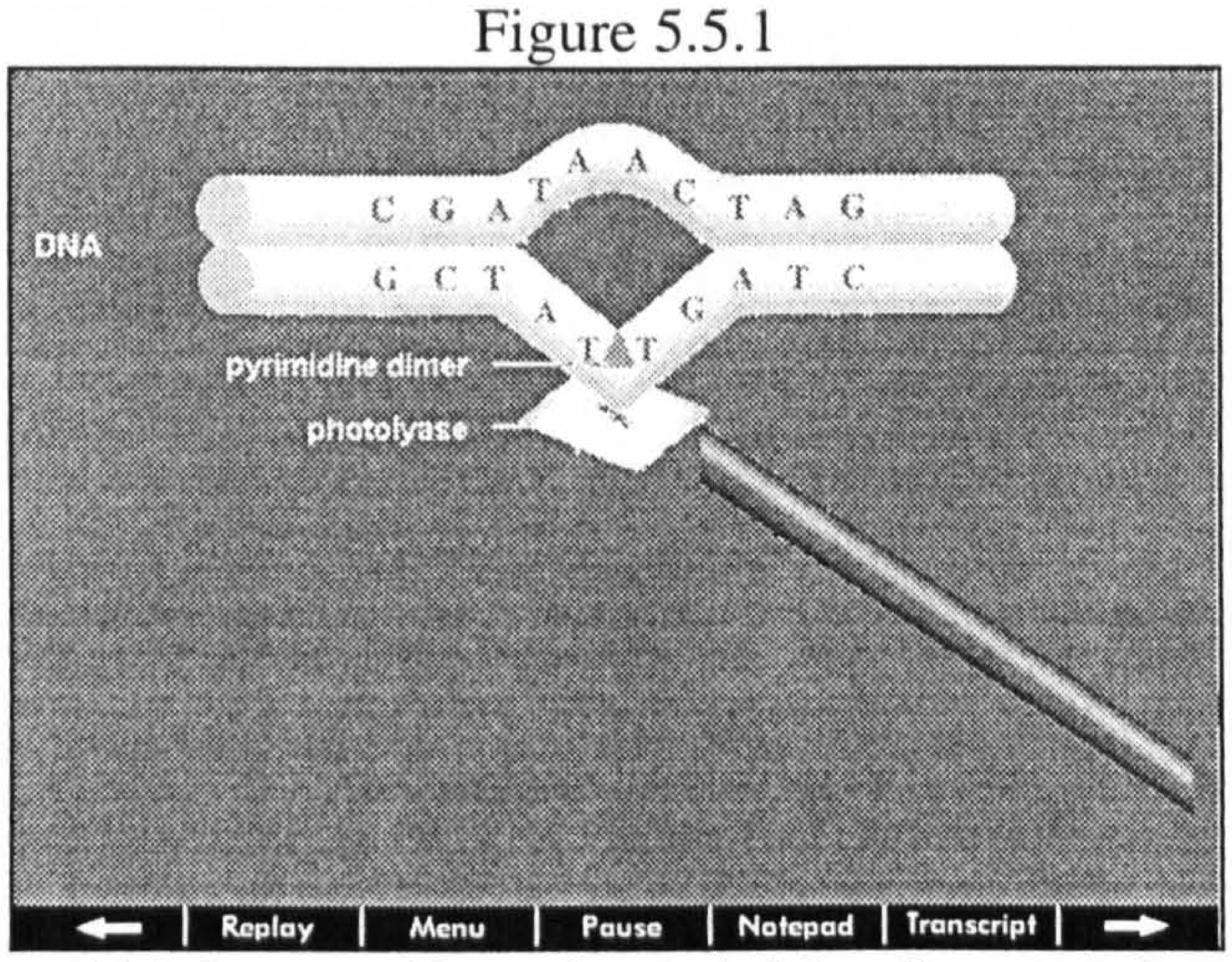
Pyrimidine dimers are reconverted to monomers ...

Figure 5.5.1



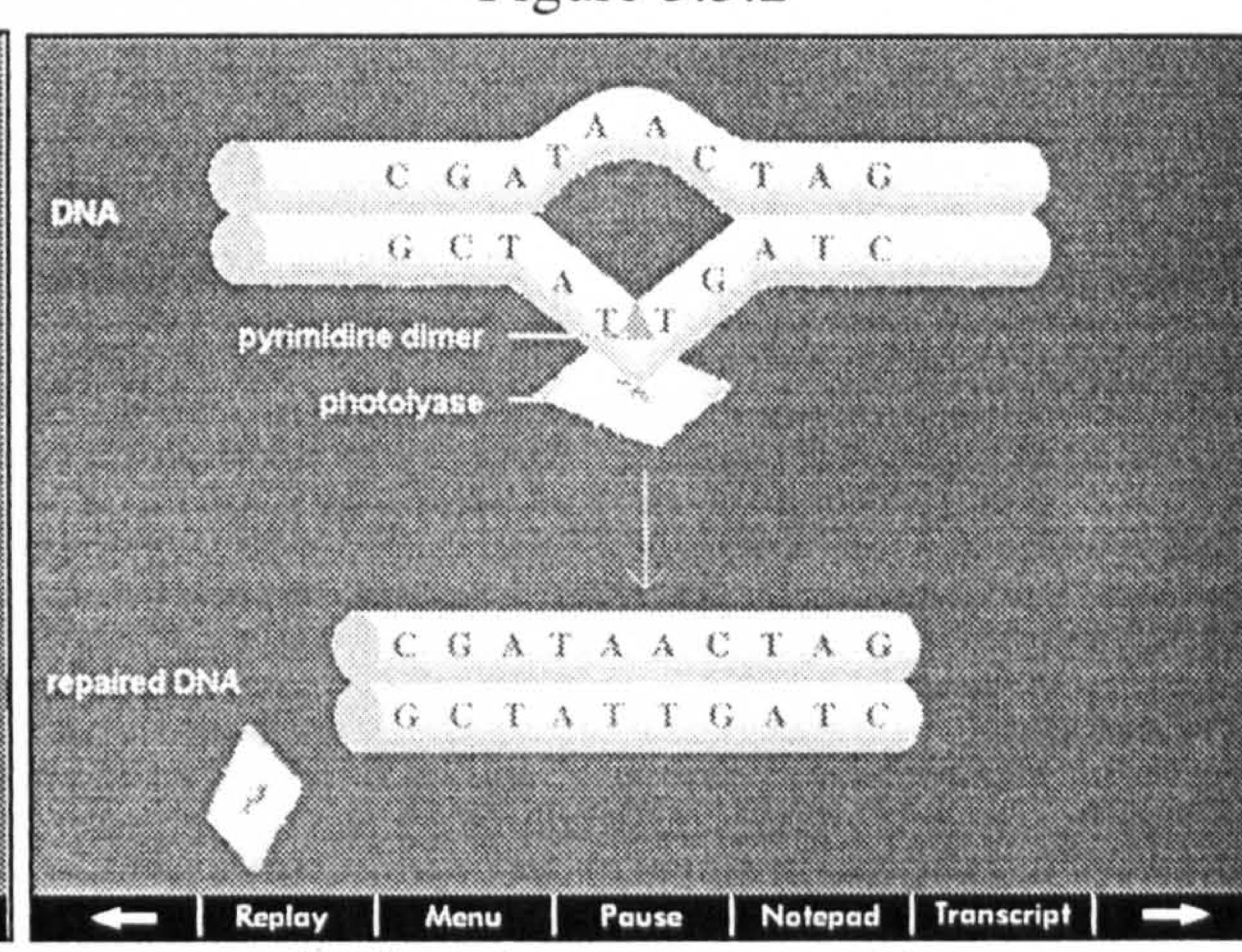
by the action of photo- reactivating enzymes called photolyases ...

Figure 5.5.2



which form a stable protien-pyrimidine dimer complex and then use light energy absorbed from the visible range

Figure 5.5.3



to convert the dimer into monomers without disrupting the double strands.

Figure 5.5.4

The photoreactivation presentation was divided into a set of ten propositions (figure 5.5.5). At the top level is a cause-effect relationship, that DNA strands are distorted by the pyrimidine dimer (shown in the centre of the DNA) and are repaired by a process called photoreactivation. This is made up of two actions. First the photolyase attaches to the pyrimidine dimer. Then, the photolyase forms a protien-pyrimidine complex with the pyrimidine dimer.

The effect of this is photoreactivation. This is made up of a sequence in which the DNA is repaired without disrupting its strands. This is made up of three actions. First the dimer protien-pyrimidine complex is struck by light energy. Then the pyrimidine dimer converts into a monomer. Finally the photolyase detaches.

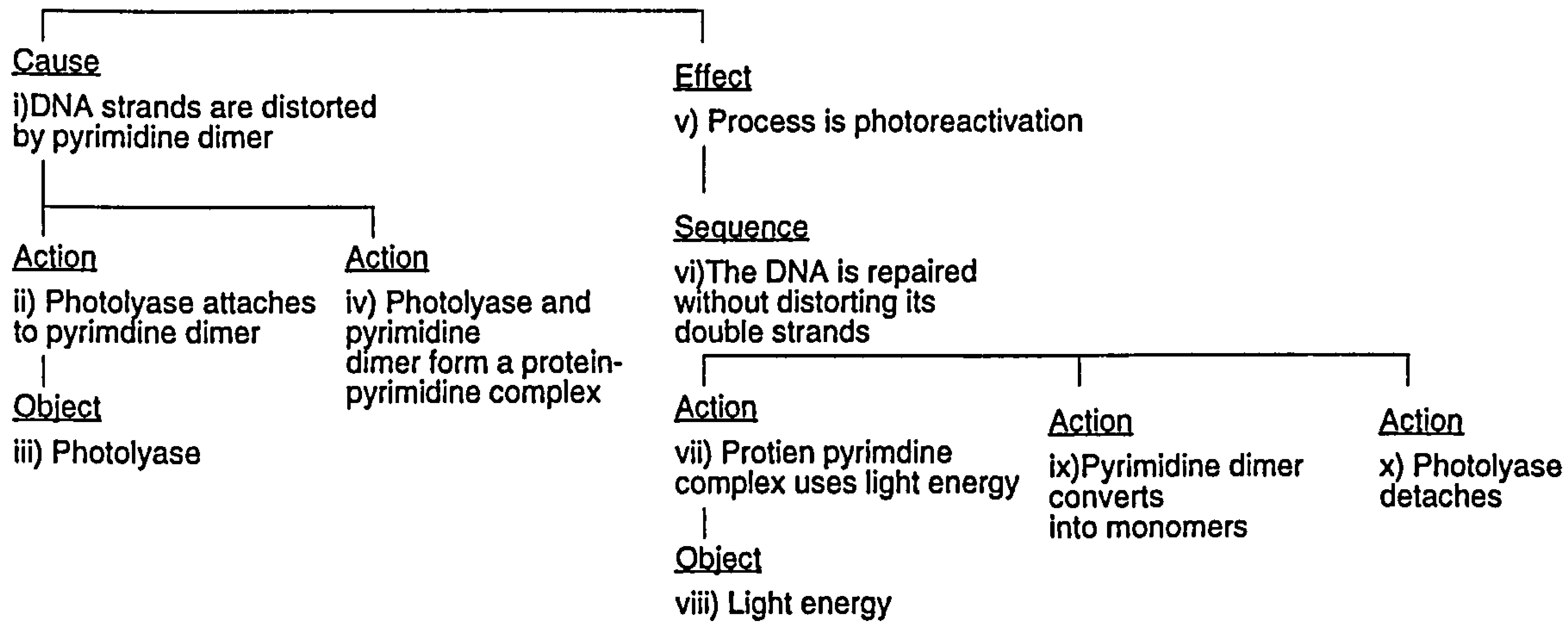


Figure 5.5.1

5.5.2 Results for 'Photoreactivation' sequence Condition 1 & 2

Proposition type	Proposition	Condition 1 Low domain knowledge (/8)	Condition 2 High domain knowledge (/8)	Total (/16)
Cause	i) DNA strands are distorted by pyrimidine dimer.	3	3	6
Action	ii) Photolyase attaches to the pyrimidine dimer.	3	7	10
Object	iii) Photolyase	7	8	15
Action	iv) Photolyase and pyrimidine dimer form a protein-pyrimidine complex	2	4	6
Effect	v) Process is photoreactivation	1	2	3
Sequence	vi) The DNA is repaired without distorting its double strands	5	7	12
Action	vii) Protein-pyrimidine dimer complex uses light energy	1	3	4
Object	viii) Light energy	6	8	14
Action	ix) Pyrimidine dimer converts into monomers	7	8	15
Action	x) The photolyase detaches	1	2	3
	TOTAL	34	52	

Table 5.5.2

5.5.3 Discussion for 'Photoreactivation' sequence Condition 1 & 2

The results are shown in table 5.5.2. There was no significant difference in recall between high and low domain knowledge subjects. The results (see table 1) show that some parts of the presentation were well recalled, whilst others were lost. Slightly more was recalled by the high domain knowledge group, in particular proposition (ii), 'Photolyase attaches to the pyrimidine dimer'. This may be because these subjects were better able to interpret the implications of the animated actions. Propositions (iii, vi-ix) were well recalled by both groups, possibly because they were conveyed by animation and reinforced by speech; while the other propositions were not. This demonstrates how poor design of the viewing sequence may lead important parts of the message (e.g the reaction in v, vi) to be missed.

Factors in the design may have effected comprehension :

- Propositions given only in the image or animation without speech cueing were poorly recalled. The initial state of the DNA, i) as damaged by the dimer was poorly recalled. This was given implicitly in the presentation by the image of the DNA shown, but was not mentioned in the speech track. Animations which were not re-inforced by speech were also poorly recalled. The photolyase attaching to the pyrimidine dimer and the photolyase detaching, x) were both shown only in animation without any speech cues. It is vital that speech describing object or actions should coincide with the object or action being shown.

- Propositions given in speech which were re-inforced by labels were well recalled. The photolyase object iii), the repaired DNA vi), and the Pyrimidine Dimer ix) were presented in labels and were well recalled. Captions or labels may be useful in re-inforceing the speech track.

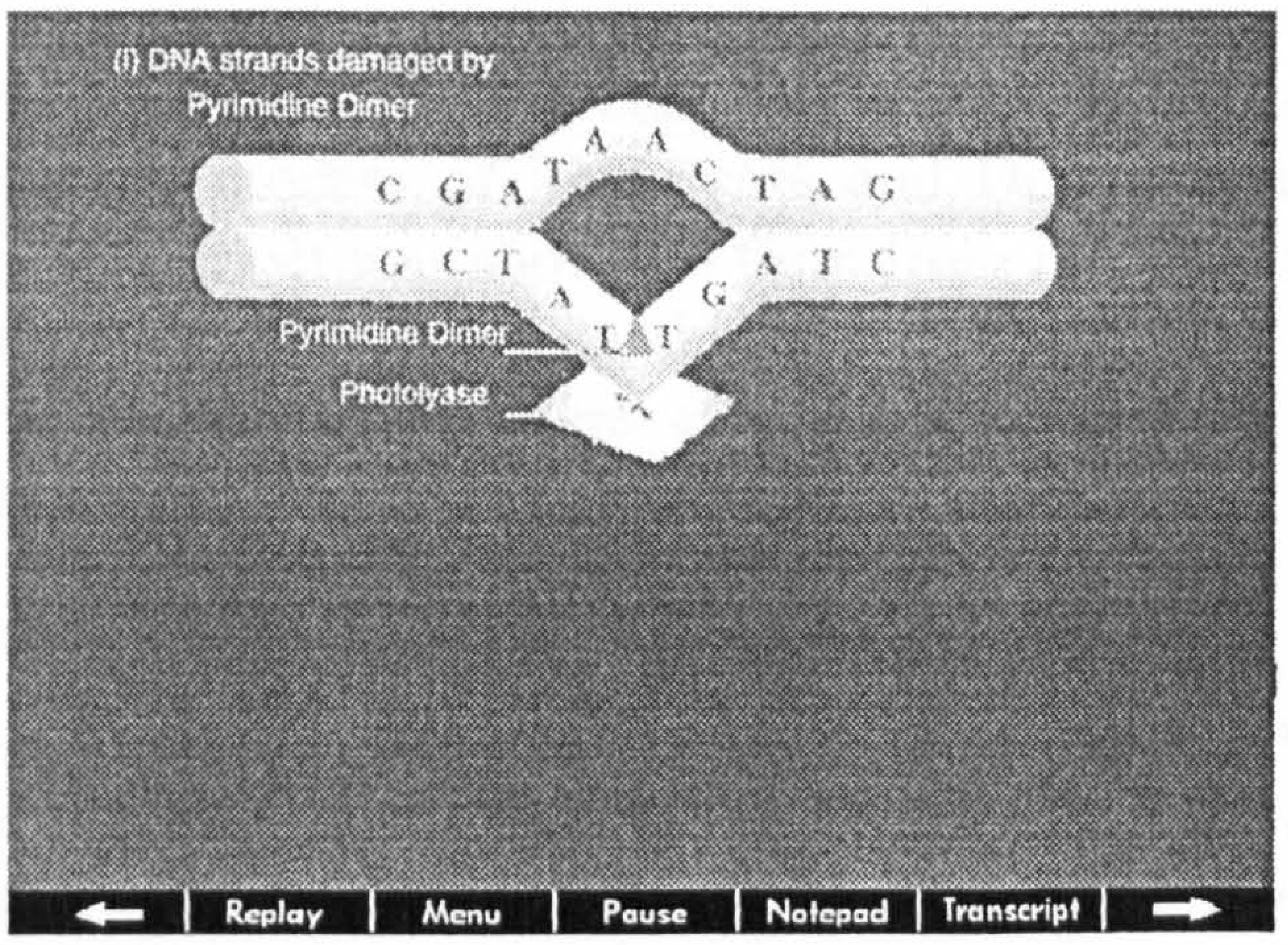
- Propositions given only in the speech track were generally poorly recalled. The propositions v) photoreactivation; and iv), vii) protien-pyrimidine complex propositions were all poorly recalled. Generally, if the speech information is complex or important then it may be best to also present it concurrently in a caption. This would seem to be particularly important for those involving the Protien-Pyrimidine complex iv), vii) which are the central propositions in the sequence.

- Objects in animations which were cued by the speech track were well recalled. Almost all subjects recalled light energy, viii), which was shown as an animation and cued with speech; and that the dimer reconverted to a monomer, ix) which was shown as a change animation between the two DNA objects and cued with speech.

- Complex actions shown in animations were poorly recalled. In particular subjects poorly remembered the animations showing formation of the protein-pyrimidine complex iv), and in the action of light energy upon the protien-pyrimidine complex vii). This may have been because these animations failed to show the effect of actions. Animation is only valuable if it shows the process taking place between objects. In the presentation, the effect of the process upon the objects was poorly shown. Thus for the formation of the protein-pyrimidine complex the photolyase and dimer should have shown the outcome, as a new object being created. For light energy striking the

Protein-Pyrimidine complex, a reaction should have been shown, indicating what that the complex was being struck. Many subjects incorrectly recalled that light energy struck the Photolyase or Dimer, or that the light energy came from the DNA.

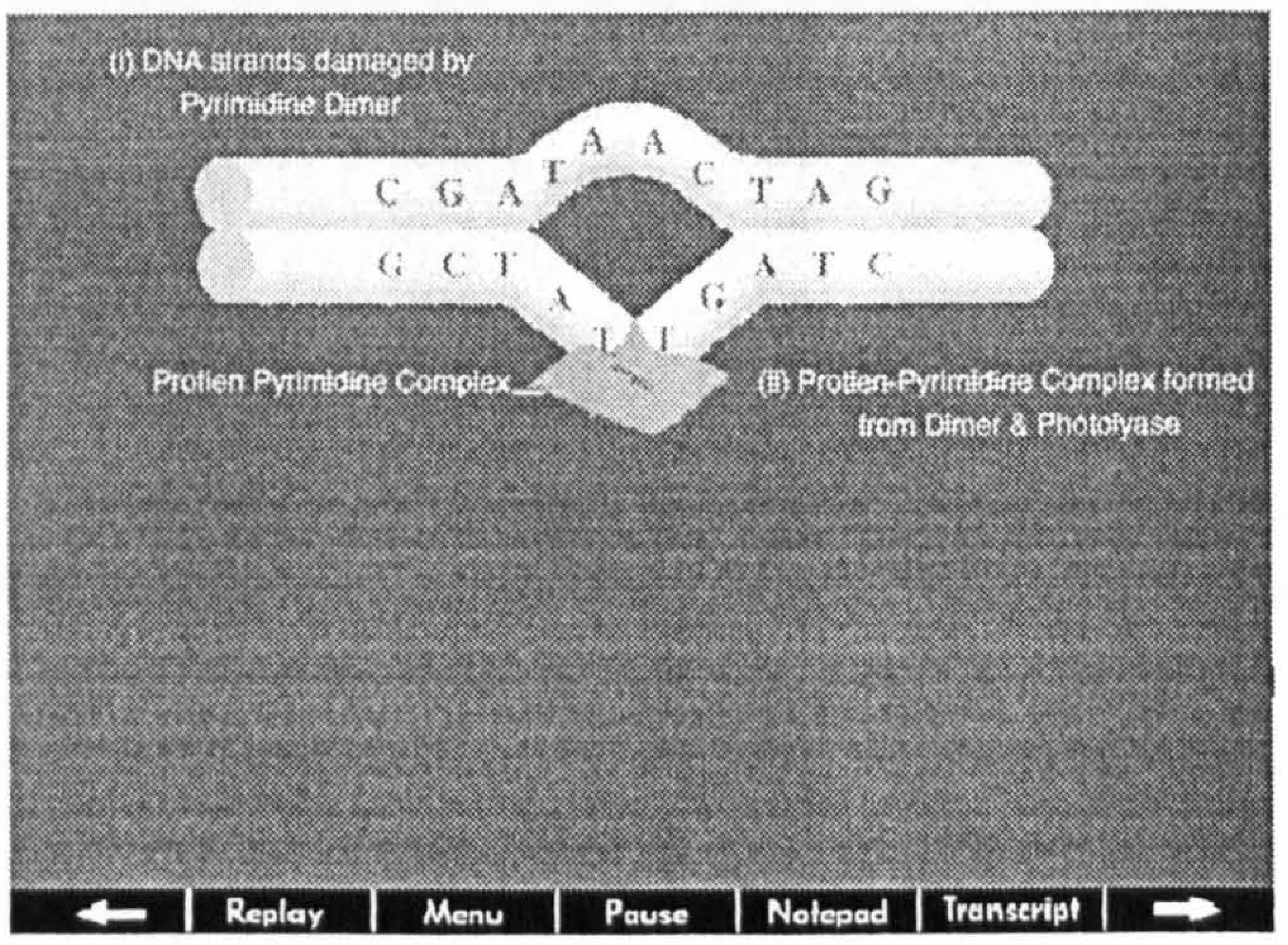
5.5.4 Re-authoring 'Photoreactivation' sequence for Condition 3



DNA is damaged by a pyrimidine dimer, which distorts the DNA's double strands. A photoreactivating enzyme, photolyase, binds to the pyrimidine dimer...

Figure 5.5.2

The redesign began with the addition of a caption to the first frame, to support the initial state of the DNA as being damaged by the Pyrimidine Dimer (fig 5.5.2) (SR6). The existing Pyrimidine Dimer label was revealed first to attract attention to Dimer object (IA5), then after allowing reading time (TA2), the new caption was revealed. A cue was also added to the speech track to support the caption (VR5).



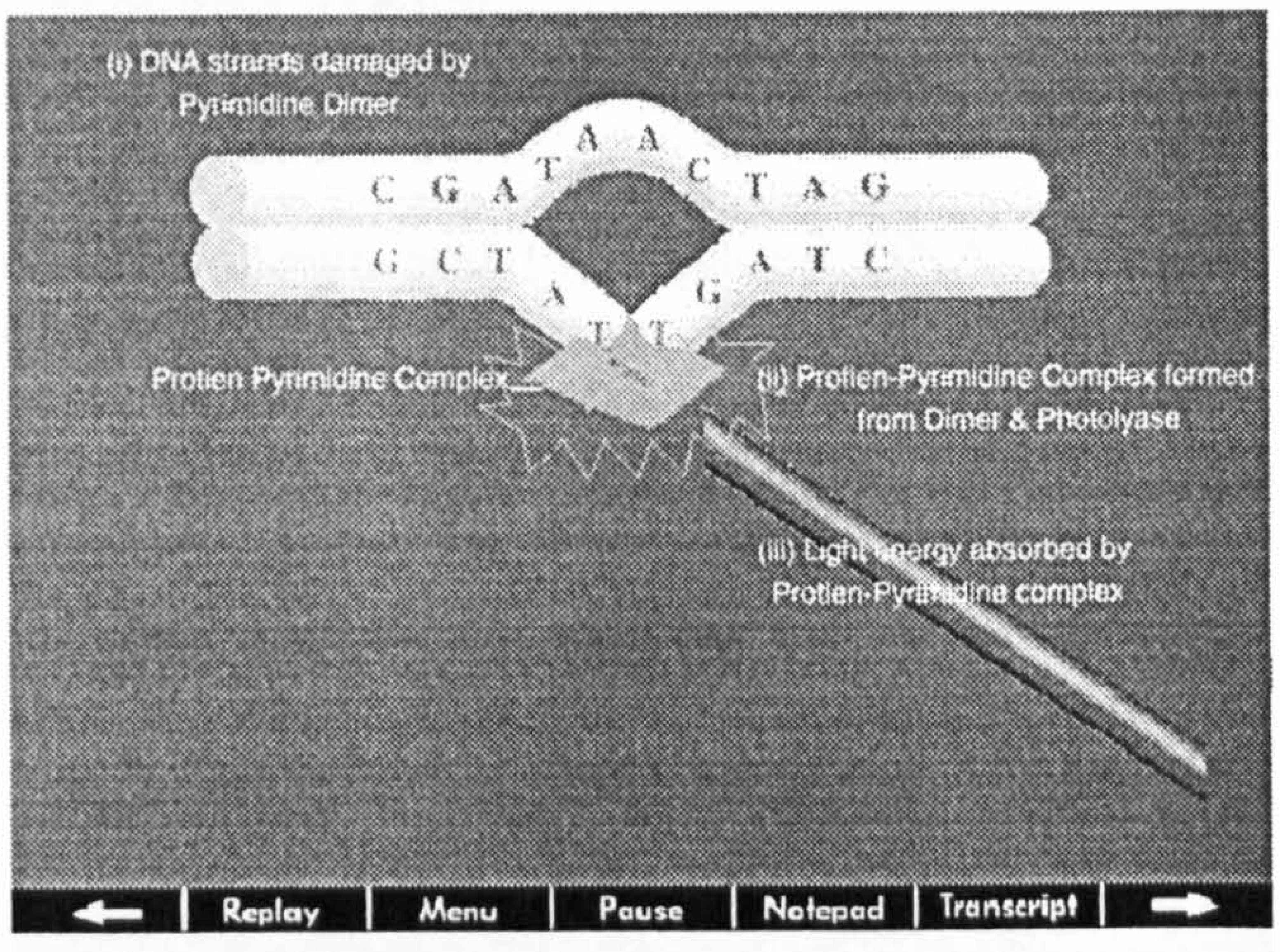
.. The pyrimidine dimer and photolyase now form a protien-pyrimidine complex...

Figure 5.5.3

The animation of the photolyase and revealed labels for Pyrimidine Dimer and Photolyase were left unchanged. To re-inforce the animation of Photolyase binding the Pyrimidine dimer, an additional speech cue was added and synchronised with the arrival of the Photolyase at the Dimer : 'photolyase binds to the pyrimidine dimer' (CP3), (VR4).

To improve recall of the Protien-Pyrimidine complex being formed, an additional animation was added to emphasise the process of Photolyase and Dimer combining (SR5). A new Protien-Pyrimidine complex object was shown to form from its two constituents, which then changed colour and shape, to signify a new combined state (fig 5.5.3). An additional Protien-Pyrimidine complex label and caption were also added to aid identification (SR1) and synchronised to be revealed with the speech track (CP3) .

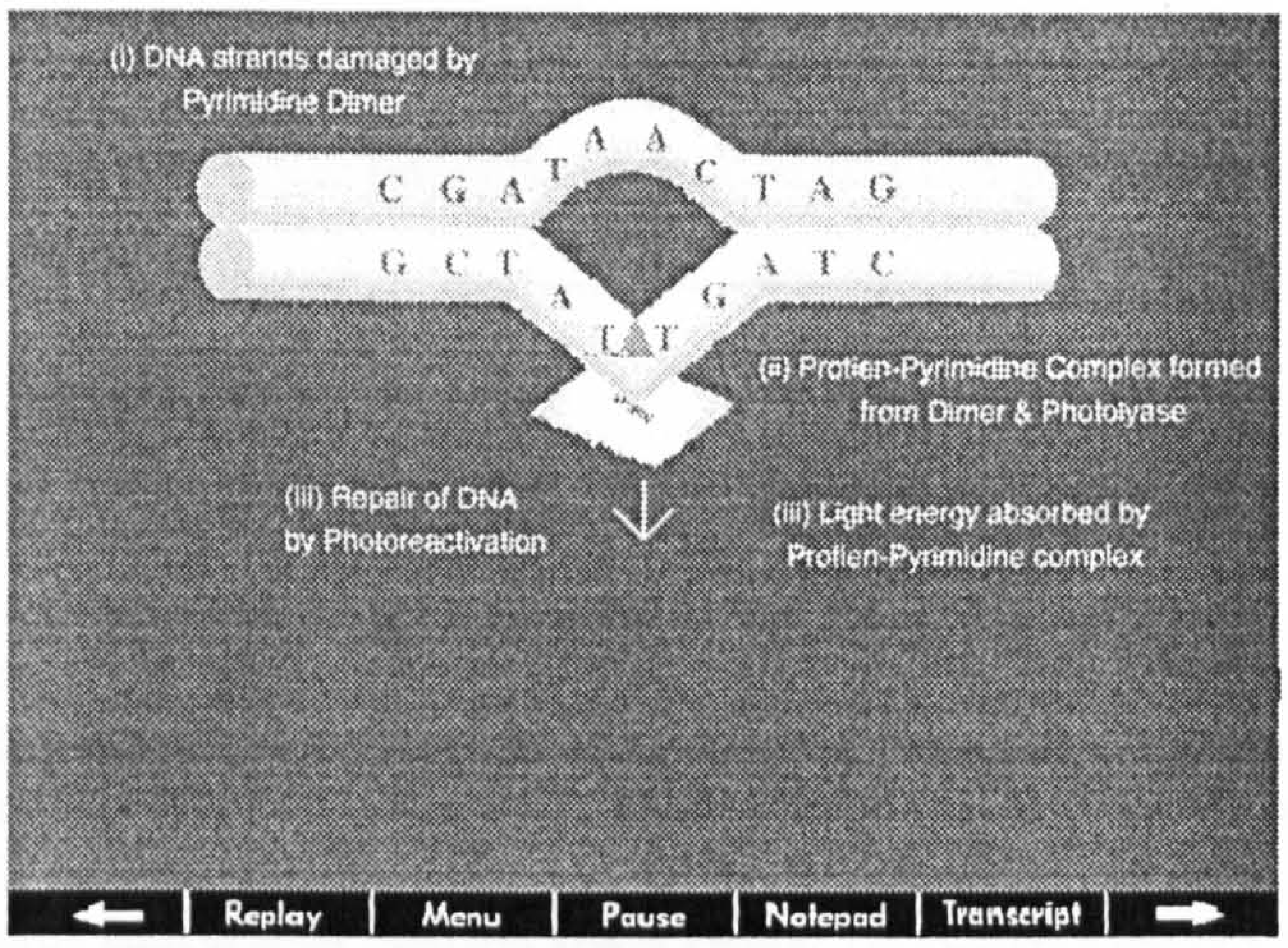
The animation of light energy striking the Protien-Pyrimidine complex was then amended, to show a reaction to the light energy strikes within the complex (SR5), with a small explosion style effect added to the complex as the light energy struck (fig 5.5.4). The light energy was paused as it struck, to allow attention to shift to it (AA3). An additional caption was shown, re-inforcing the speech (VR3) .



Next, protein pyrimidine complex absorbs
light energy from visible range.

Figure 5.5.4

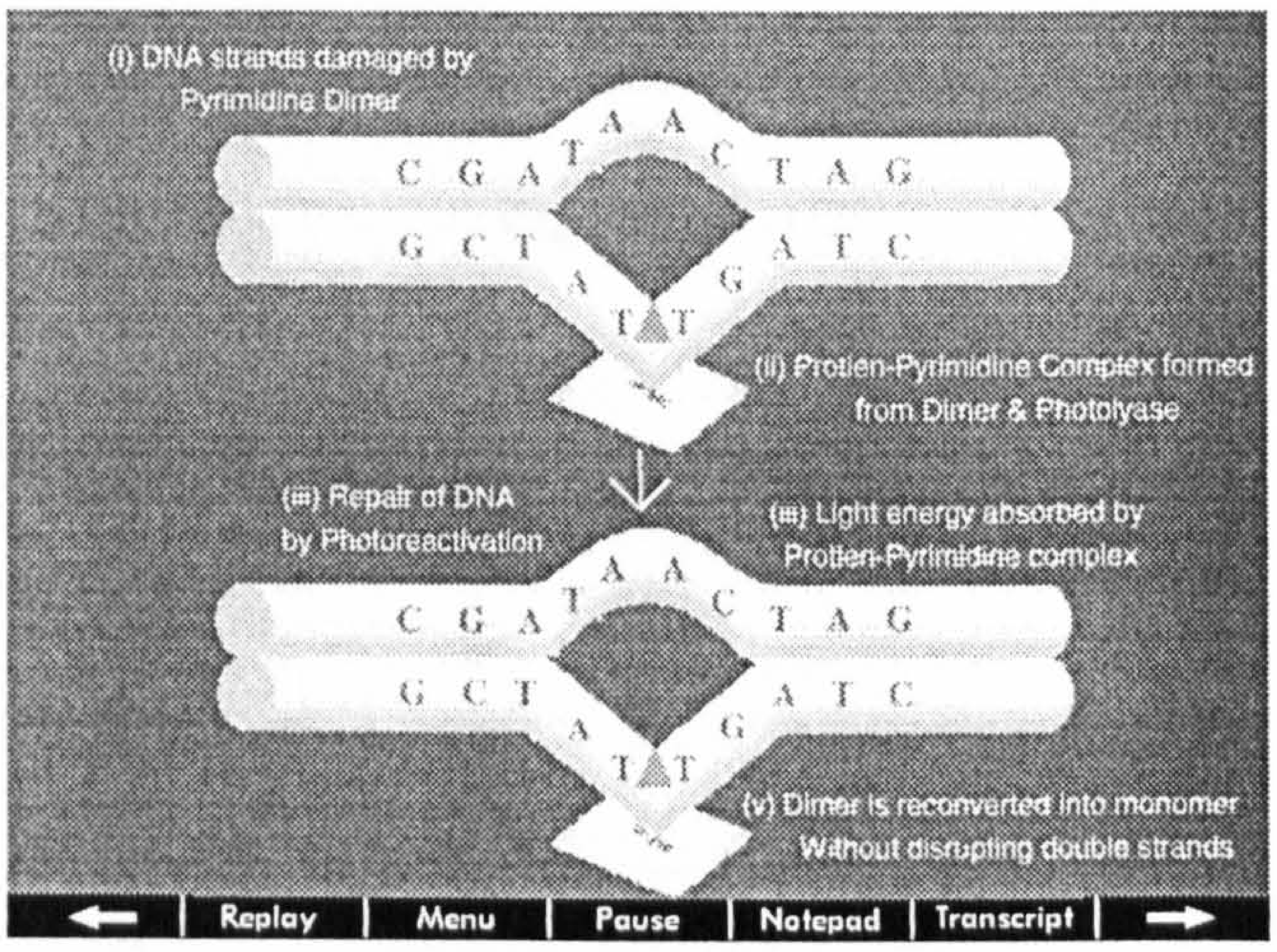
After the light energy animation was finished, the Protein-Pyrimidine complex was shown to pull apart, separating the photolyase and dimer (fig 5.5.5). An additional caption and speech cue re-inforced 'photoreactivation' (VR5) .



... This causes the DNA to be repaired by photoreactivation...

Figure 5.5.5

The arrow symbol was then shown and the repaired DNA revealed. To make the process clear, an animation showed the dimer disappearing and the DNA re-joining (fig 5.5.6) (SR5).



The pyrimidine dimer is now reconverted to a monomer, without disrupting the DNA double strands.Finally, the photolyase dissociates from the DNA.

Figure 5.5.6

Finally, the repaired DNA label was given as a caption and moved to the centre of the DNA, where attention had been directed by the animation (fig 5.5.6). Reading time was then allowed (TA2). The animation of the photolyase was shown, with an added speech cue 'photolayse detaches' to re-inforce it (VR4).

5.5.5 Re-authoring text & speech only Condition 4

The speech & text only condition was based on the speech track from the re-authored presentation. It was given as single paragraph describing the photoreactivation process.

DNA is damaged by a pyrimidine dimer, which distorts the DNA's double strands. A photoreactivating enzyme, photolyase, binds to the pyrimidine dimer. The pyrimidine dimer and photolyase now form a protien-pyrimidine complex. Next, the protien pyrimidine complex absorbs light energy from the visble range. This causes the DNA to be repaired by photoreactivation. The pyrimidine dimer is now reconverted to a monomer, without disrupting the DNA double strands. Finally, the photolyase dissociates from the DNA.

Figure 5.5.7

5.5.6 Results for reauthored and speech only 'Photoreactivation' sequence Condition 3 & 4

Proposition type	Proposition	Condition 3 Re-authored Multimedia (/8)	Condition 4 Speech & Text (/8)
Cause	i) DNA strands are distorted by pyrimidine dimer.	7	3
Action	ii) Photolyase attaches to the pyrimdine dimer.	6	2
Object	iii) Photolyase	7	5
Action	iv) Photolyase and pyrimidine dimer form a protein-pyrimidine complex	5	2
Effect	v) Process is photoreactivation	4	1
Sequence	vi) The DNA is repaired without distorting its double strands	7	4
Action	vii) Protein-pyrimidine dimer complex uses light energy	5	2
Object	viii) Light energy	8	3
Effect	ix) Pyrimidine dimer converts into monomers	5	1
Action	x) The photolyase detaches	6	3
	TOTAL	60	26

Table 5.5.3

5.5.7 Discussion for 'Photoreactivation' sequence Condition 3 & 4

The results are shown in table 5.5.3 in the 'Multimedia' column. A significant difference was found between recall in the low domain knowledge group in the original design and the improved design (Mann-Whitney U, $P<0.05$). There was a clear improvement in recall over the original design, and in some cases performance was superior to those of the high domain knowledge group, although this difference was not significant. The following improvements are of particular note :

- Recall of the initial state of the DNA, i) was improved by the addition of a caption and speech cue to the original visuals.

- Recall of the simple animation photolyase attaches, ii) and detaches, x) were both improved by the use of speech cues to the animations.

- Recall of photoreaction, v) improved by using a caption and redesigning the speech track to be make this object more salient.

- Recall of the more complex animations involving the Protein-Pyrimidine complex being formed, iv) and being struck by the light energy were both improved, vii). These animations were re-authored to clarify the process, and reinforced by an additional label and caption.

Recall performance with text-speech was significantly worse than for the low knowledge group in the improved multimedia design (Mann Whitney U, $P<0.001$), but was not significantly different from the low knowledge group for the original presentation. This again showed that for re-authoring to provide additional linguistic information was not in itself sufficient to improve comprehension : animation and images were also vital for low knowledge subjects to understand the presentation.

5.6 Conclusions

The studies of subject's recall showed that design problems may lead to difficulties in comprehension. We suggested comprehension may be improved by design of components for attracting the users' attention, thereby ensuring that the presentation could be formed into coherent whole. This may be achieved by scripting the speech track to more tightly reference and synchronise with the visual elements of the presentation; and by using animations and captions to support the content.

		Total Recall		
Authoring Condition	**Domain knowledge**	**Metastatis**	**Ames test**	**Photoreactivation**
Condition 1 : Original	Low domain	40	35	34
Condition 2 : Original	High domain	42	50	52
Condition 3 : Re-authored	Low domain	65	52	60
Condition 4: Speech & Text only	Low domain	37	24	26

Table 5.6.1 Total recall scores for each condition

	Significance (Mann-W)		
Authoring Condition	**Metastatis**	**Ames test**	**Photoreactivation**
Condition 1 : Original Low domain 'v' Condition 2 : Original High domain	No sig	No sig	No sig
Condition 1 : Original Low domain 'v' Condition 3 : Re-authored Low domain	<0.001	No sig	<0.05
Condition 2 : Original High domain 'v' Condition 3 : Re-authored Low domain	<0.05	No sig	No sig
Condition 3 : Re-authored Low domain 'v' Condition 4 : Speech & Text only Low domain	<0.001	<0.001	<0.001
Condition 1 : Original Low domain 'v' Condition 4 : Speech & Text only Low domain	No sig	No sig	No sig

Table 5.6.2 Mann-Whitney test results for each condition

Media Type change applied to	Total Design Changes Metastatis	Ames test	Photoreactivation
Text (change original text, add new text)	11	6	7
Image (add image, change original image, add symbol)	4	3	2
Speech (change original speech, add new speech)	7	5	7
Animation (change original animation, add new animation)	4	2	4
Total :	26	16	20

Table 5.6.3 Total number of design changes made (addition or change to a particular media type)

Comparing across the results tables (see table 5.6.1, 5.6.2, 5.6.3) for the three sequences, several conclusions can be made :

- For the original presentation, domain knowledge seemed to help little in recall. Total recall scores were generally poor for all of the original sequences, and there was no significant difference found in any of the sequences between high and low domain subject groups. This was surprising, since the high domain knowledge subjects were familiar with complex biochemistry materials. However, a poor presentation may have prevented them from being able to use their knowledge.

- Re-authoring improved total recall scores above those for low domain subjects shown the original presentation. There was a significant difference from the metastasis and photo-reactivation sequence. The Ames test may have benefited less from re-authoring, as it had fewer original problems than the other presentations. The table of total design changes shows that the metastasis sequence had most changes made, followed by the photoreactivation sequence.

- Re-authoring also improved total recall scores above those for high domain subjects shown the original presentation. There was a significant difference only for the metastasis sequence. This suggests that problems in the original metastasis sequence must have been so severe as to prevent the high domain group benefiting from their knowledge.

- In the speech & text conditions, low domain subjects showed far worse total recall scores than for the re-authored presentation, with significant difference in all cases. This result suggests that the use of multimedia can improve recall over conventional language based media alone. Comparing speech & text recall with the original presentation provides evidence that multimedia alone is insufficient; the presentation design hence is effective.

The total recall scores were similar between text & speech and original multimedia for low domain subjects, and their was no significant difference between them.

It may be dangerous to over generalise the results of the studies for several reasons. First, the studies used a slightly unusual condition in which subject's were only allowed to view the presentation once. This may have exaggerated certain problems in the original presentations which repeated playing could have been solved.

Second, the studies only tested immediate recall. Whilst this provides a useful indicator of comprehension, long term retention may have varied. It would thus have been useful to further test recall after a period of a week to determine if the improvements in presentation design had along range effect on retention.

Third, the re-authored presentations all had longer playing times than the originals, so this may have possibly allowed a greater chance for the subjects to comprehend the material. However, the longer playing time was due to the use of specific guidelines which suggested that reading and viewing time for materials must be allowed, so such changes were inevitable if the guidelines were to be applied fairly.

Fourth, the studies made limited use of some of the media selection guidelines. This is because in many cases the correct media were present, and the design issues concerned those of within media design and attentional design.

Fifth, the studies are weakened because in applying the guidelines a large number of changes were made in the presentation. Whilst care was taken to control the content so that the propositions were the same for both presentations, some propositions were implicit within the original presentation i.e. only shown as part of an animation, whilst in the re-authored version they were given by text or speech. This poses a question of whether like against like conditions was used between the original presentation and re-authored presentation on low level issues such as attentional design, when such marked changes were made in within media design. To resolve this issue further studies would be required which differentiate the impact of each type of guideline.

Chapter 6

Validation with Novice Designers.

6.1 Introduction

This chapter reports a study on the utility of the design method. The study was performed as part of a workshop on the design method given to Philips Research. It was performed with two groups of novice designers, each accompanied by an experienced multimedia designer. The groups were given the task of using the method to produce expository presentations. The study had several aims :

- Investigate how the design method could be used by novice designers to produce multimedia presentations. Would novices could make use of each of the method steps to produce a satisfactory multimedia design ?

- Test out the utility of the method. Would the method could be applied by groups who were not connected with its development, and who had little background knowledge in multimedia design ? A facilitator in each group kept track of the methods steps used by the group and noted down any difficulties encountered in using the guidelines.

- Use professional designers to criticise the method. What would be the response of experienced designers to the method ? A concern here was whether it would conflict with their existing design knowledge, and how they would adapt to using guidelines rather than instinct.

6.2 Study Method

Two groups of designers were given a two day tutorial on the design method. The tutorials focused on the use of the method for multimedia design, and in providing skills in critiquing existing multimedia designs. This provided an overview of the cognitive issues in multimedia design (see chapter 2), and provided a detailed review of the method (chapter 3) and motivating studies (chapter 4 &5).

The groups were made up of two novice designers and one professional designer. The novice designers had existing software development skills, but had little or no experience of designing multimedia presentations. The professional designers had a background of skills in multimedia design, and had been formally trained at an art college. Both were currently employed as interaction designers and had knowledge of commercial multimedia authoring environments, such as Director. One had also worked for approximately 12 months for a multimedia design company, Voyager Interactive.

It was hoped that the professional designers would be able to apply their own knowledge to the design problem and provide criticism of any shortcomings in the method. It was also hoped that the designers could give an indication of the acceptability of guideline based advice in the design process.

The task of the groups was to produce a short expository presentation, based upon a car maintenance task. Becuase the exercise was intended to last for only half a day, only a relatively simple sequence could be tackled in the time.

The source for these materials was an existing commercially produced car maintenance manual. The original materials were largely text based, accompanied with one on two pictures. The novice designers were encouraged to review these and use the design method to critique the original presentations, and to author their own versions. One group was given the task 'How to change the transmission oil', whilst the other task was 'Adjusting the contact breaker points'.

The groups were given six hours to produce a storyboard for their presentations, with the hope that they would be able to roughly animate their work on video to give some idea of how the finished presentation would appear. The groups also had access to video equipment and digital cameras.

The groups were provided with a paper copy of the guidelines, and each group was accompanied by a facilitator who had a considerable knowledge of the design method. The facilitators did not assist in producing the designs, but provided advice on how to use the method.

6.3 Transmission oil sequence

6.3.1 Task Model

The transmission oil sequence was made up of the steps required to locate the transmission in the car's engine bay, find the dipstick and then check the oil level, topping it up if it was below a recommended level. For simplicity, the set-up procedure is not shown.

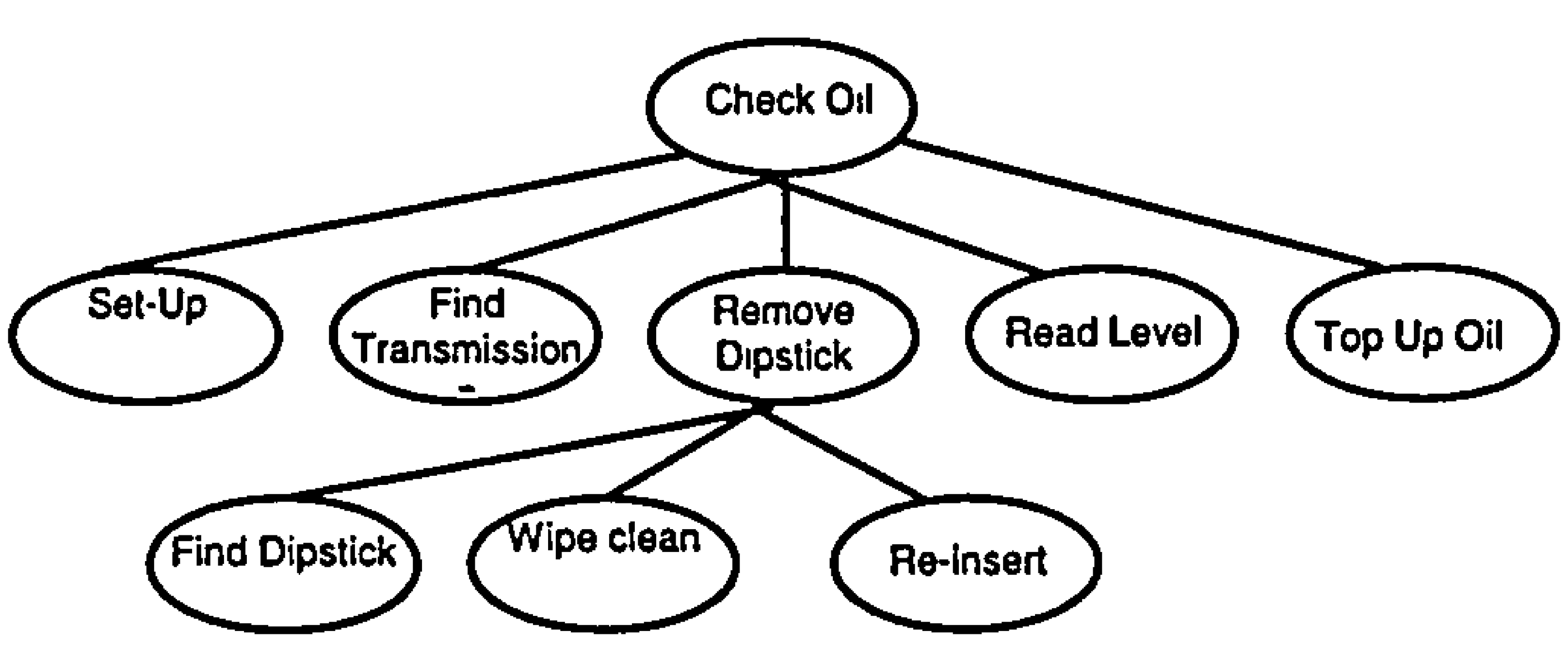


Figure 6.2.1 Gold Standard Task model for transmission oil task.

Some of the materials produced by the novice designers are shown below. Figure 6.2.2 shows an example of the task media model, which contains the procedures, objects and actions in the task; together with the media selected which are annotated at the base of the task model.

Figure 6.2.3 shows an example of the attention graph produced for part of the sequence. Each of the media given in the task-media model are set out on the y axis, whilst the task steps run across the x axis. The designers chose to organise the sequence to reveal each part of the presentation separately. Certain media were also left in focus to provide context. This is shown by the use of dotted lines in the graph.

Figure 6.2.4 shows an example of a script produced from the presentation, which was used to produce the final presentation. The novice designers felt in was useful to bring together the various media they had sequenced in the attention graph together with the wording of the speech track. References to pictures are embedded within the script, suggesting where contact points were required between the media. Figure 6.2.5 - 6.2.8 show the presentation which the subjects produced.

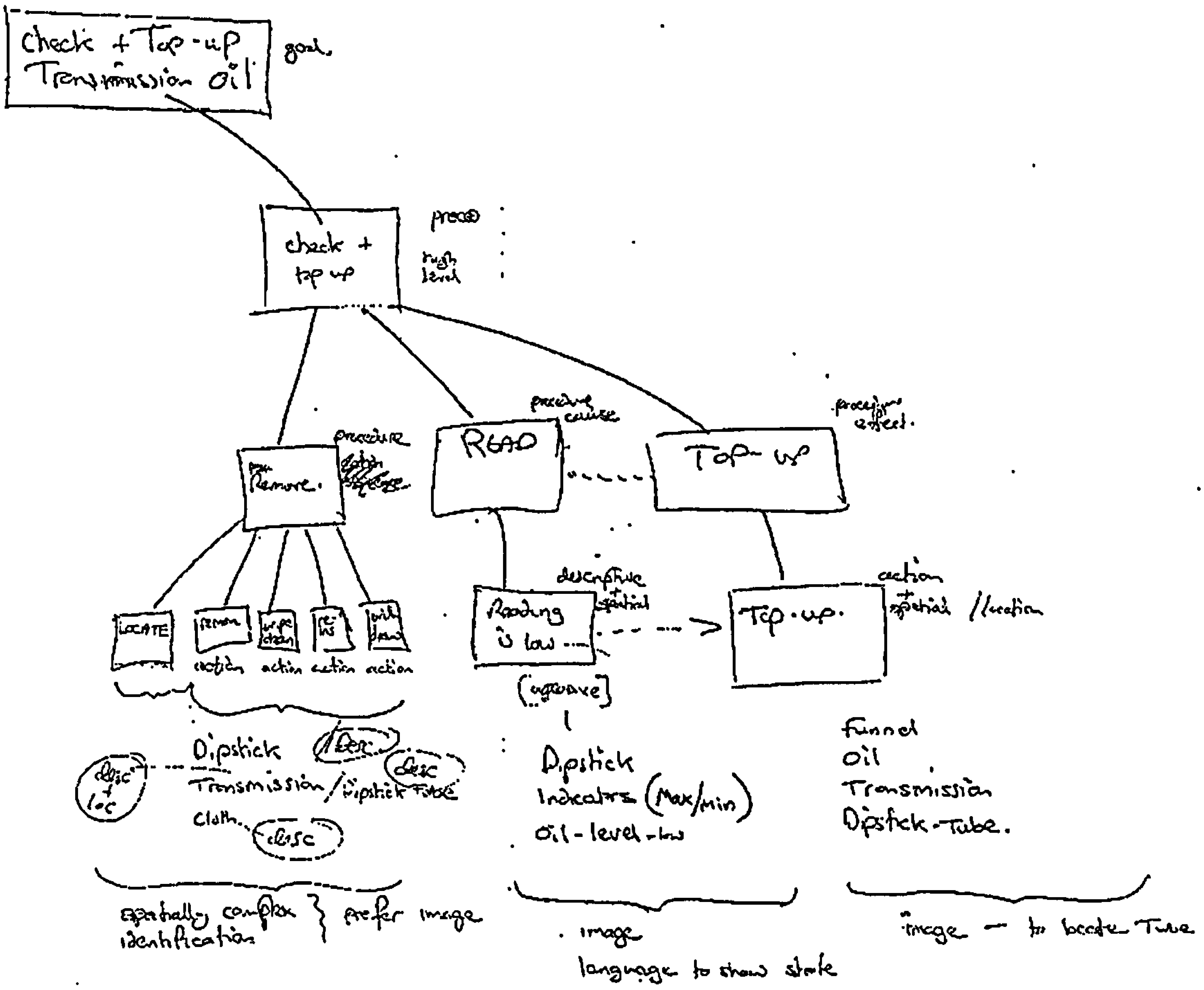


Figure 6.2.2 Task-Media model for transmission oil task.

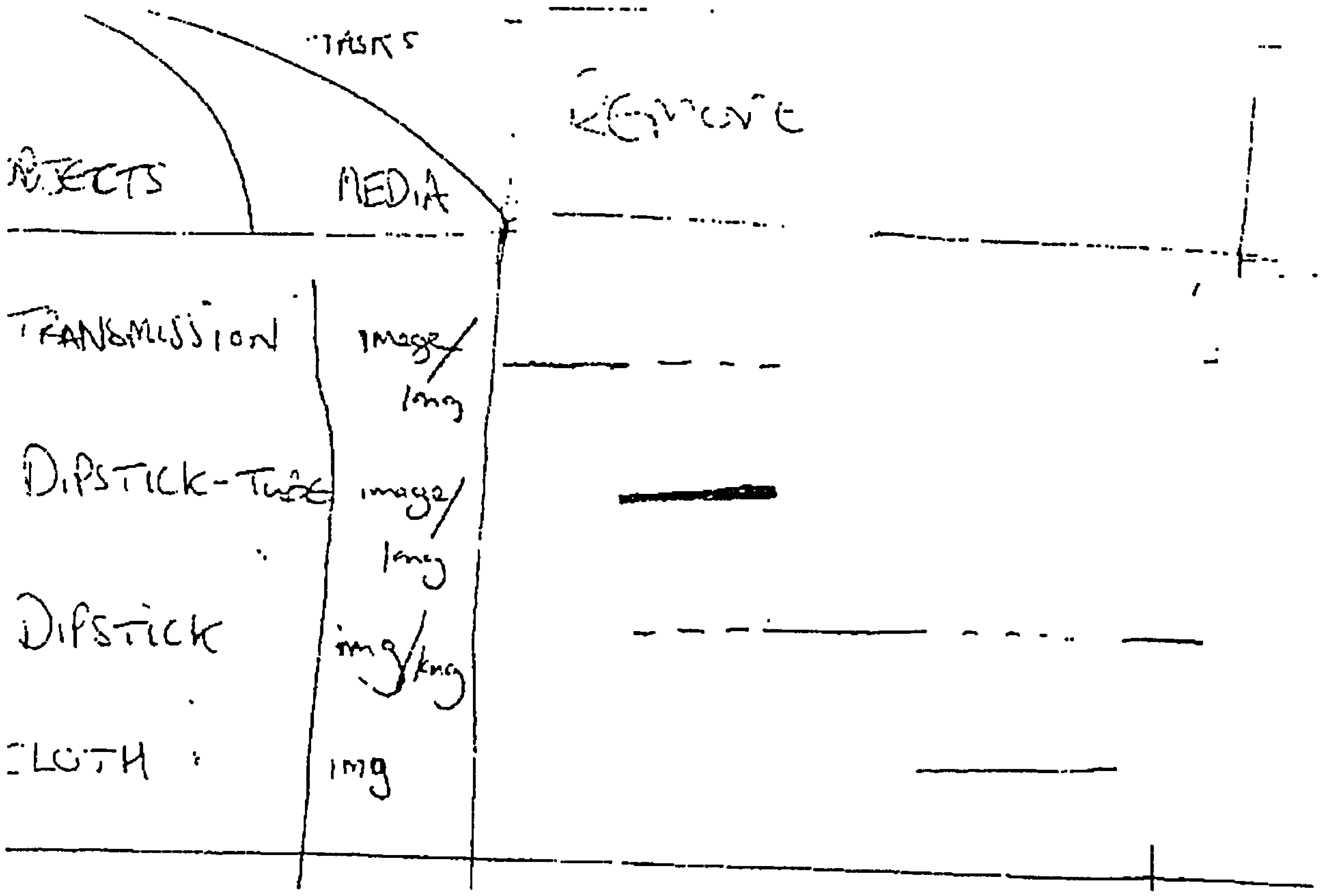


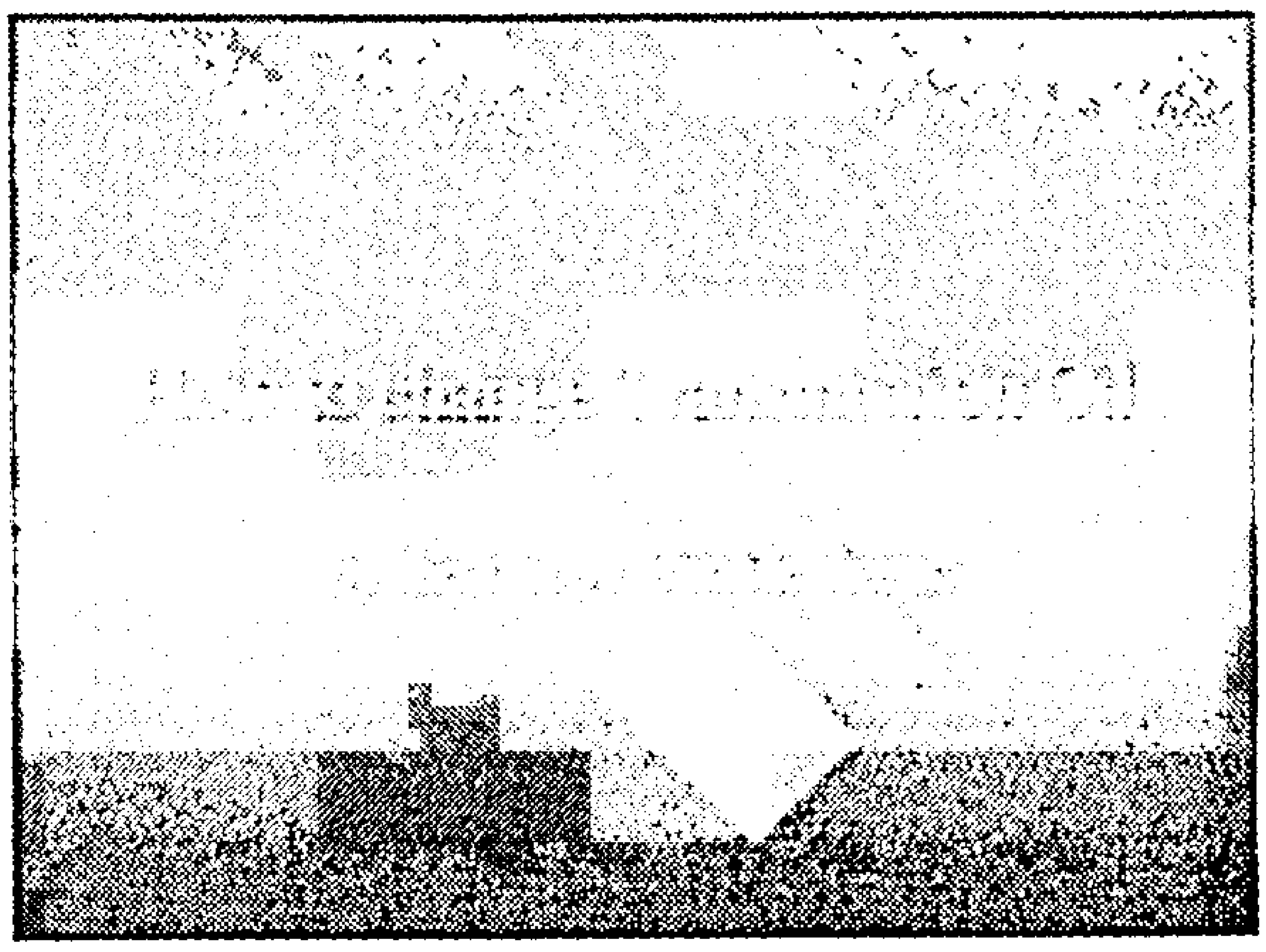
Figure 6.2.3 Attention graph for 'remove dip stick' step of transmission oil task.

- ~~Now~~ ~~then~~ Then, re-insert the dipstick for a few seconds
  <<Cut to static image/graphic of dipstick>>
- The fluid should be between the minimum and maximum indicators

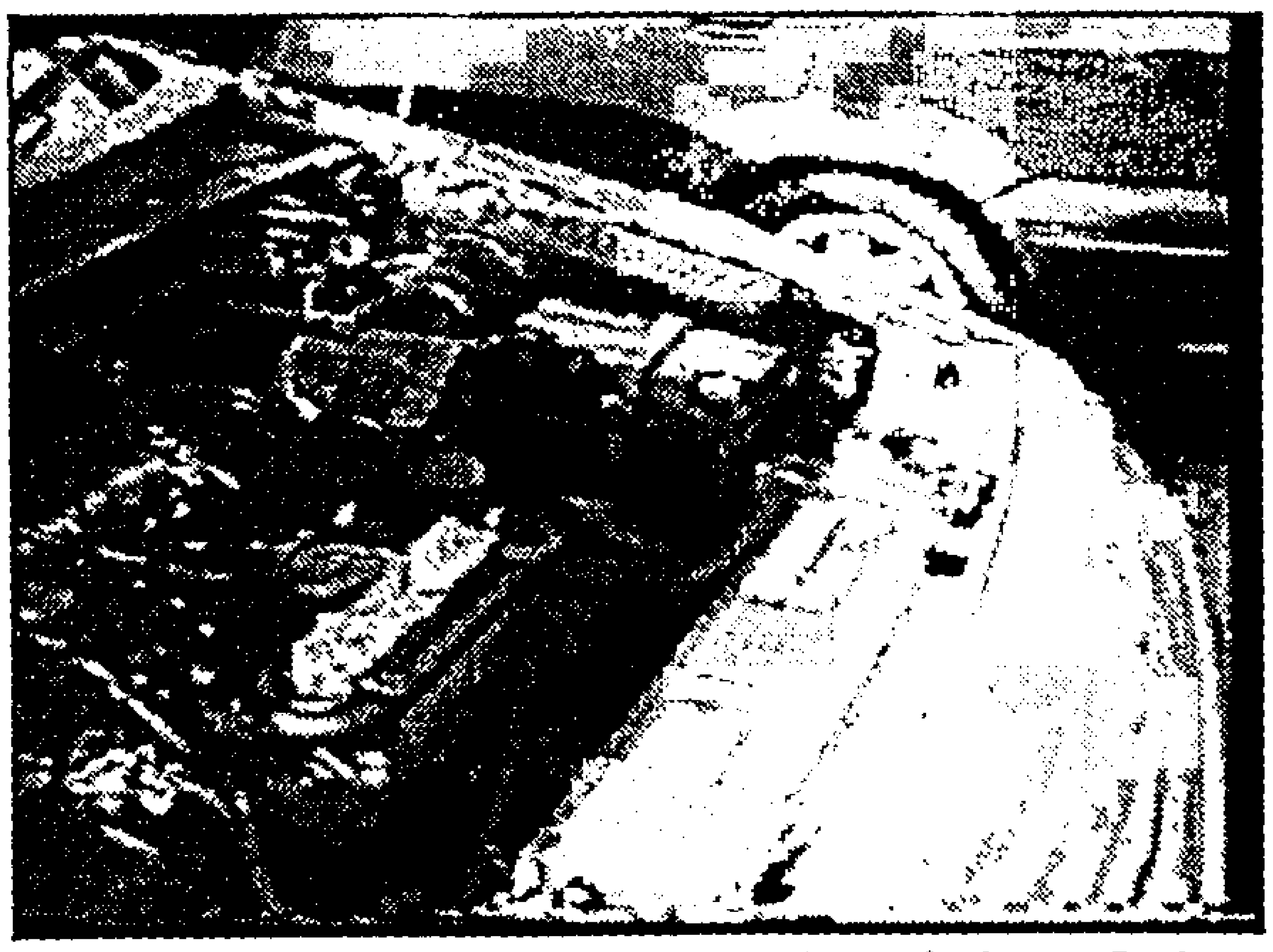
4. If the oil level is low then top up with transmission fluid

Figure 6.2.4 Script for transmission oil task 're-insert dipstick' step

6.3.2 Presentation design



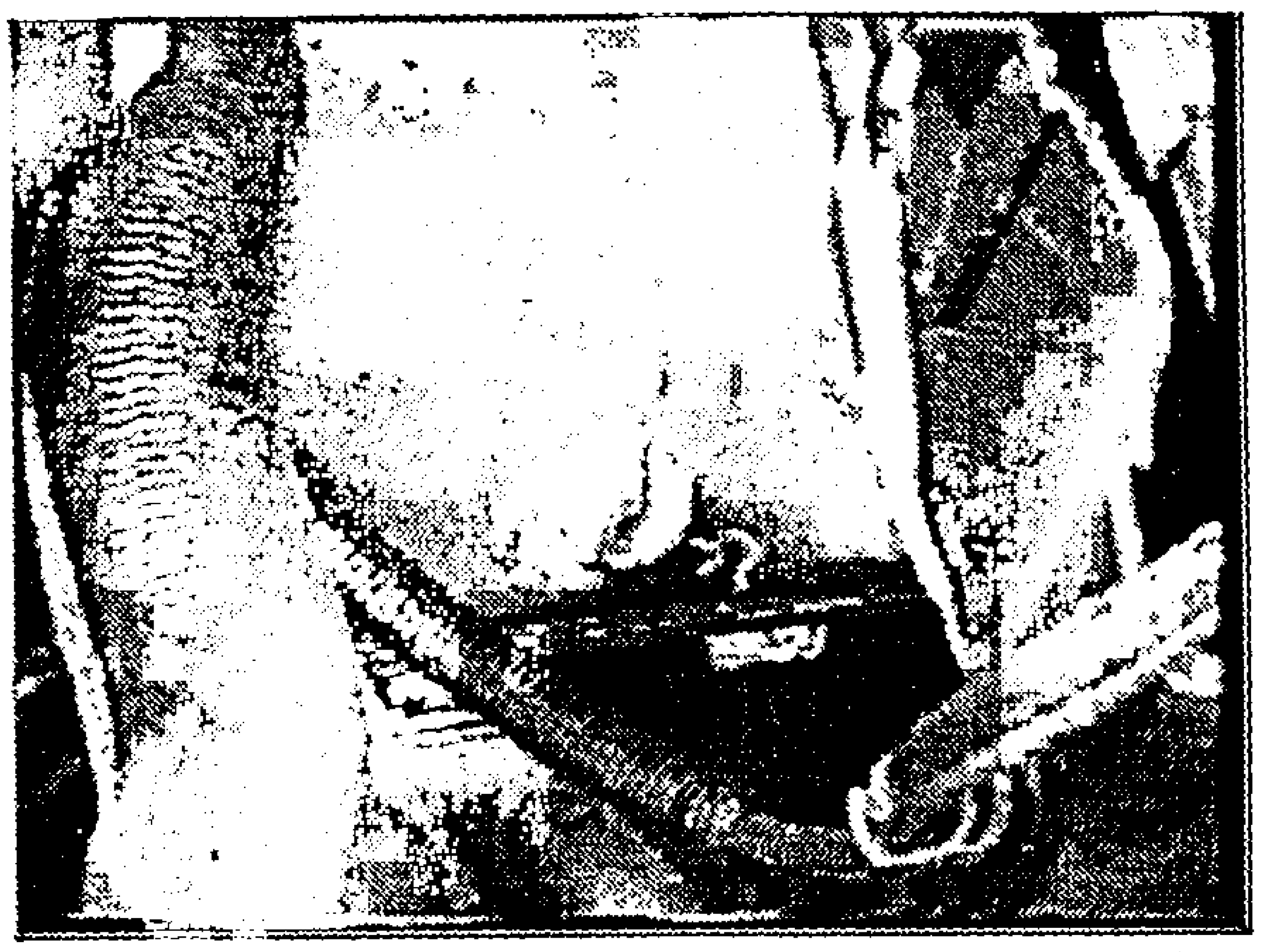
How to change the transmission oil.



The transmission lies to the right of the engine.

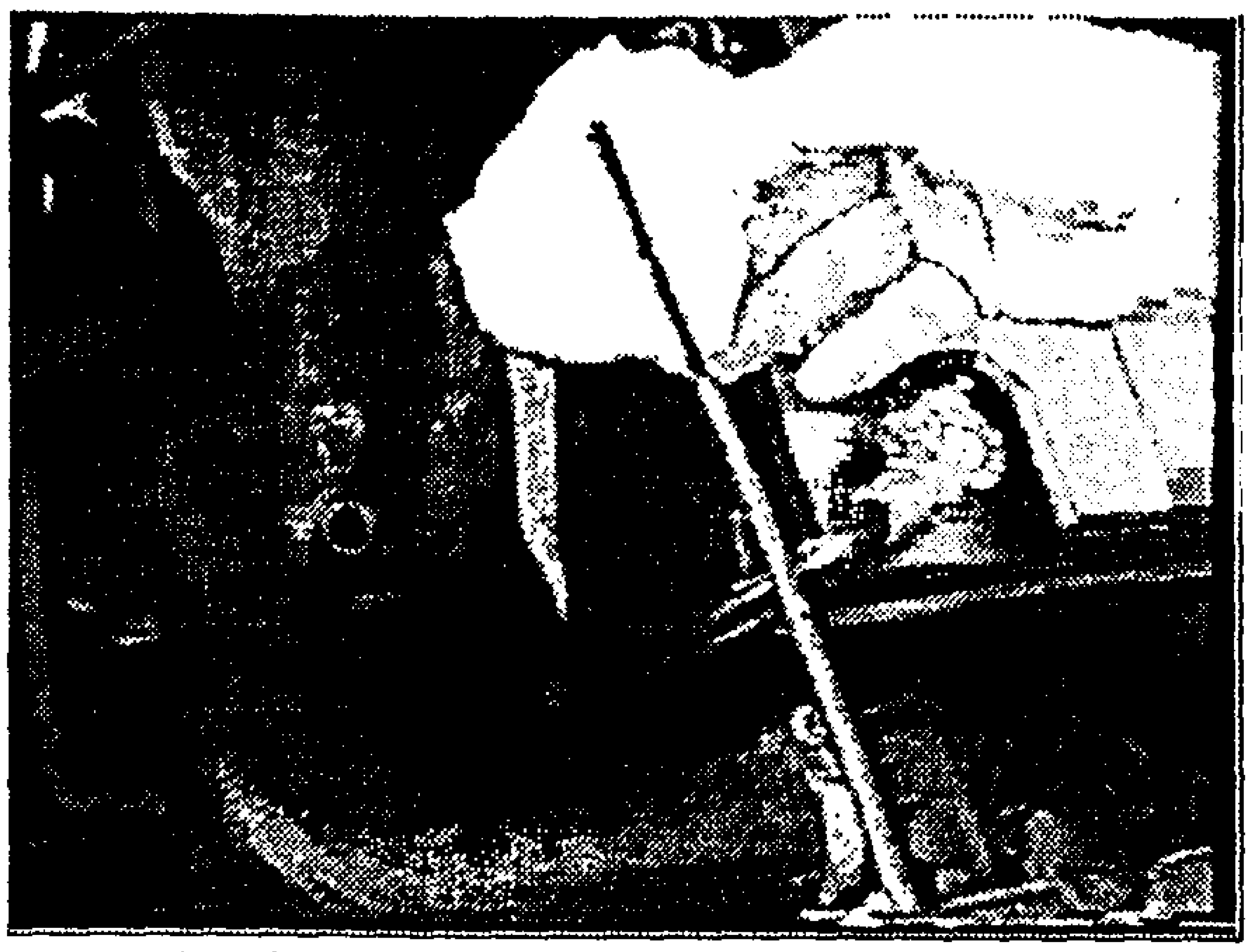
Figure 6.2.4 Presentation sequence for find transmission step

You can see the orange dip stick at the front of the transmission.

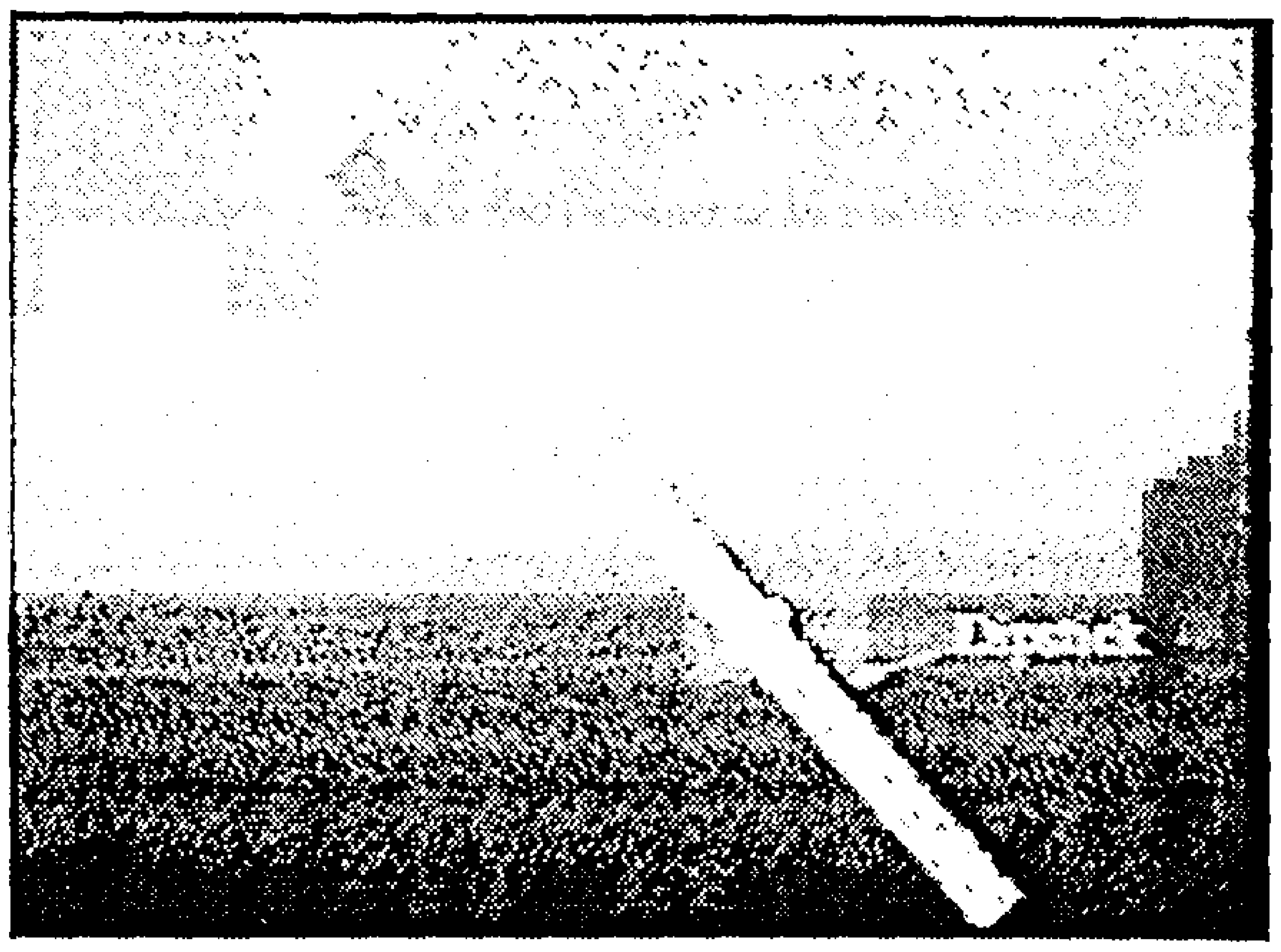


Now remove the dipstick from the dipstick tube...

Figure 6.2.5 Presentation sequence for remove dipstick step

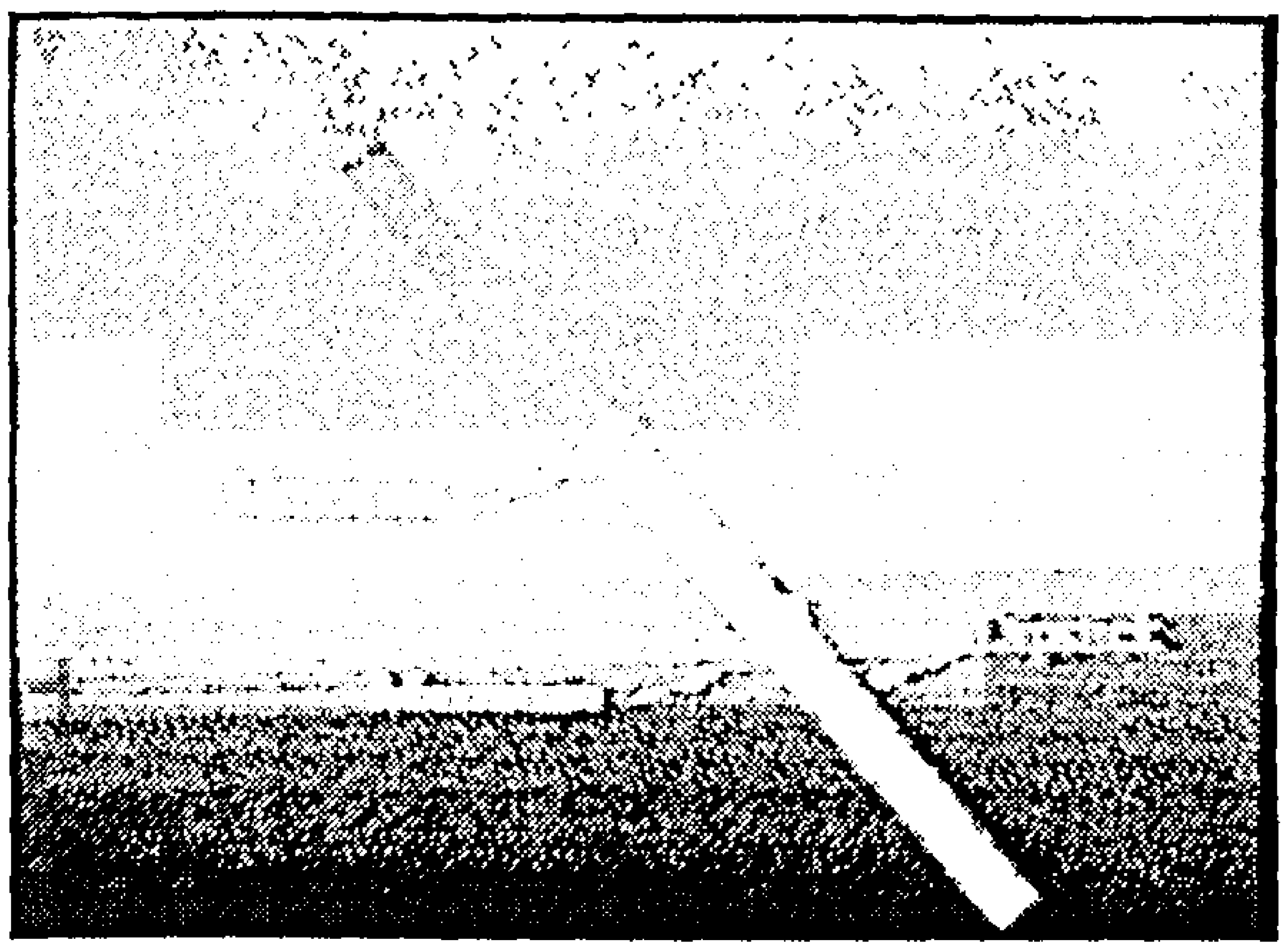


...and wipe it clean. Then re-insert it for a few seconds.

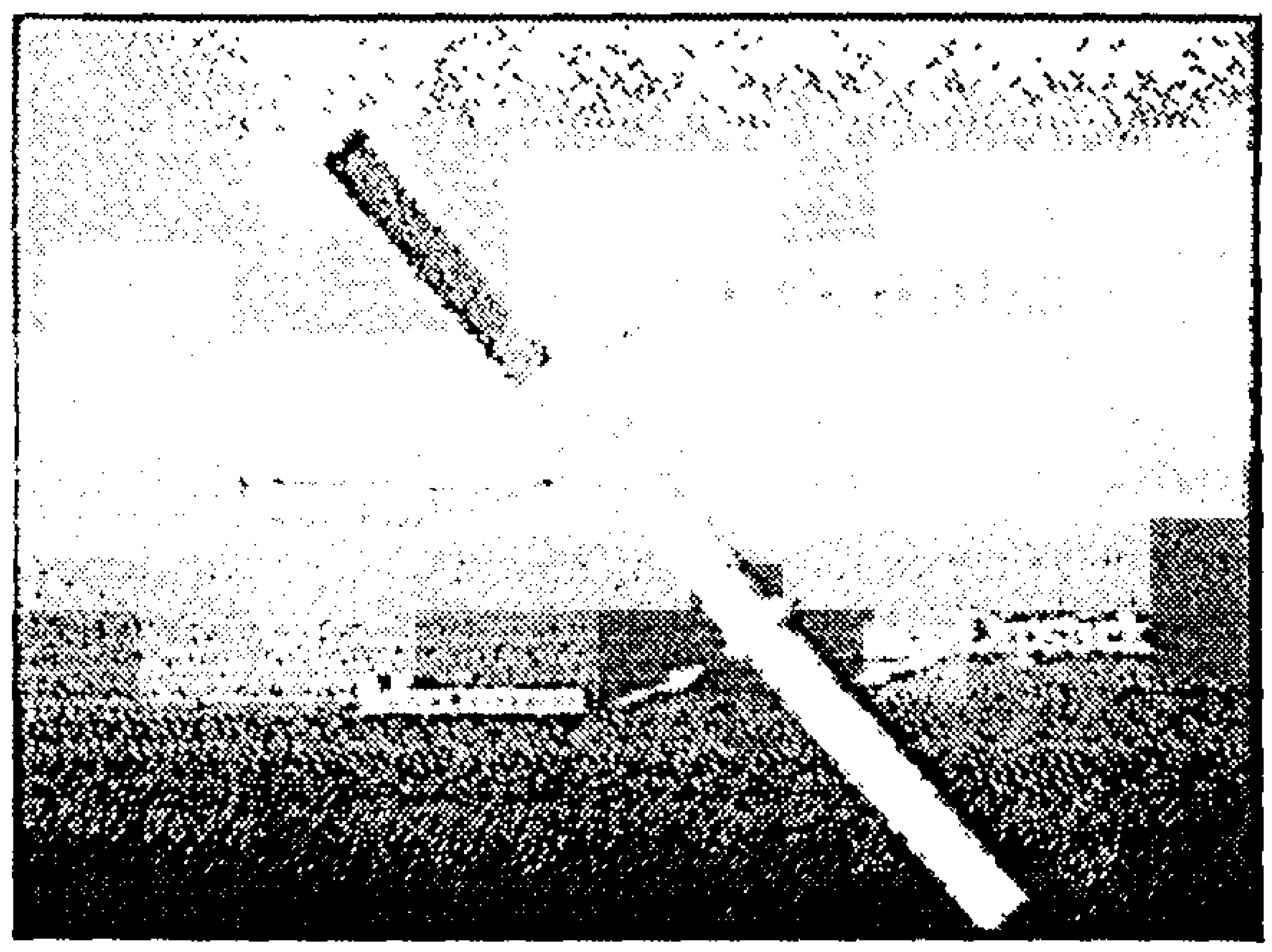


The oil level on the dipstick...

Figure 6.2.6 Presentation sequence for re-insert dipstick step

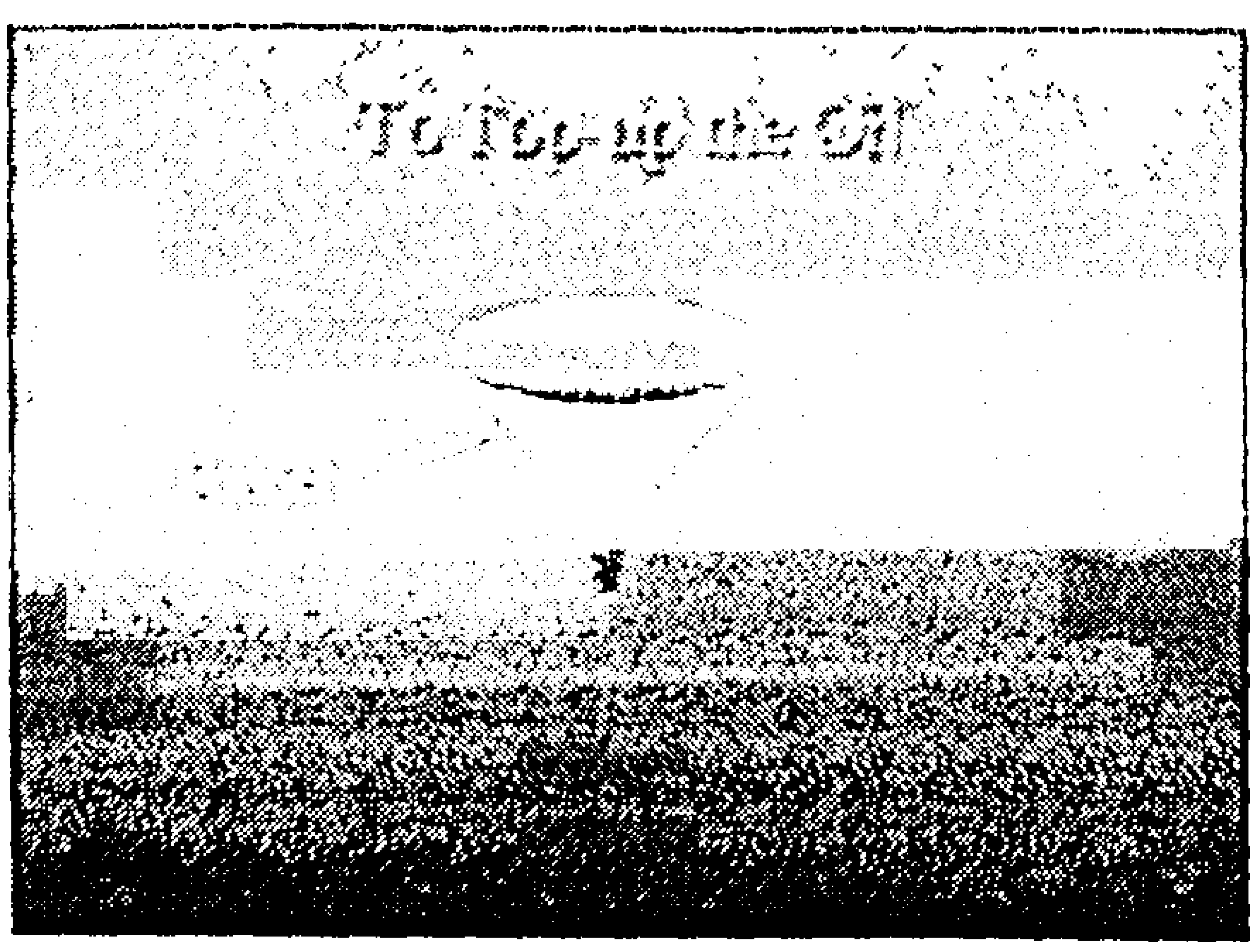


...should be between the minimum and max indicators.

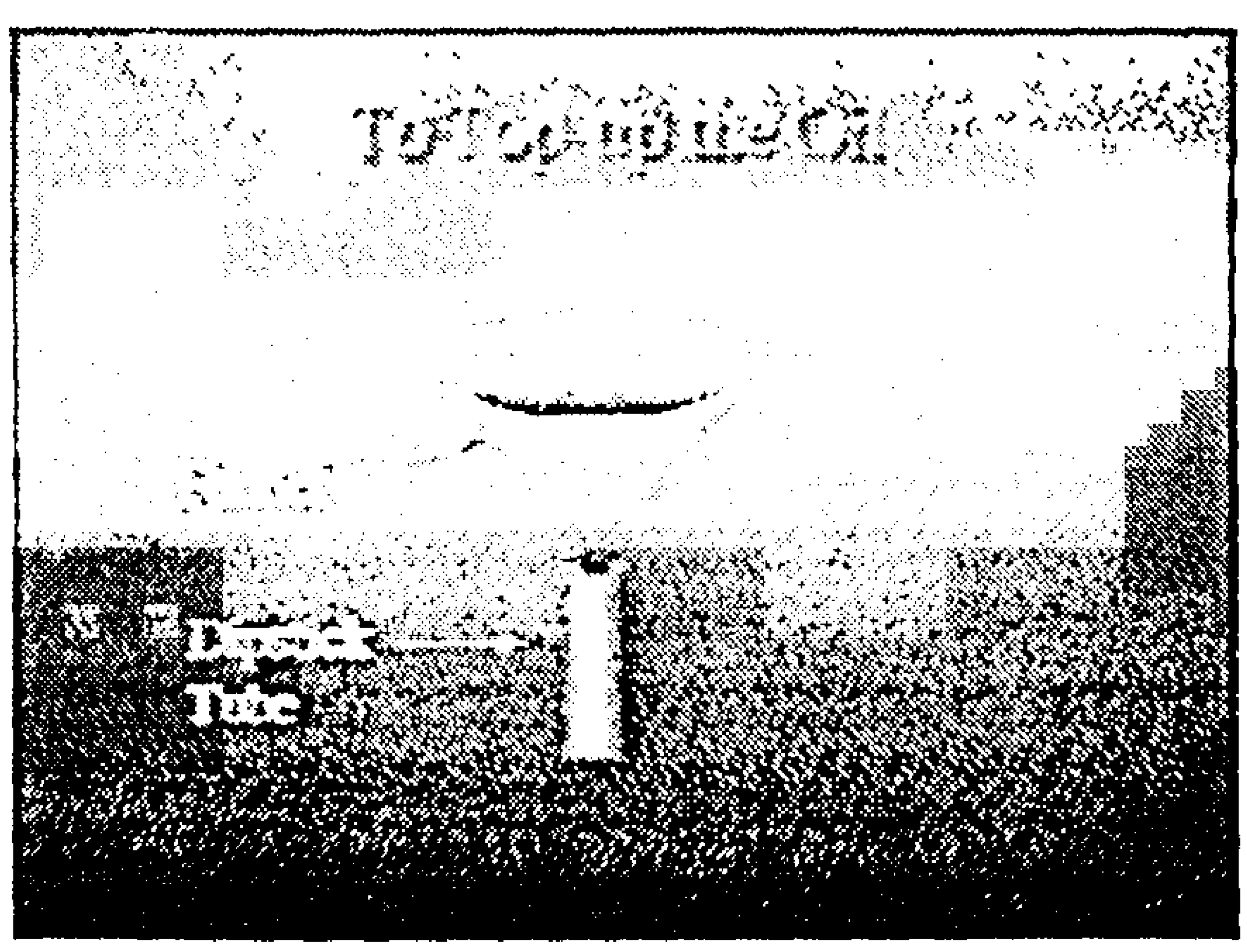


If the oil level is low then top up with suitable transmission oil.

Figure 6.2.7 Presentation sequence for check oil level step



To top up the oil place a funnel...



into the dipstick tube and add oil.

Figure 6.2.8 Presentation sequence for top up oil level step

6.3.3 Discussion

- The guidelines were used successfully to choose appropriate media. The presentation produced by the group used a better integrated mix of visuals and language than the original car maintenance materials. When the novice designers stepped through the task model and applied the media selection rules, they found that the spatial, descriptive and operational nature of many of the task steps called for a mix of image and language (SR1, SR3, SR5).

- Whilst the designers had few problems in choosing information types, they did find that the process was repetitious, requiring that very similar decisions were required repeatedly. They complained that this meant re-consulting the guidelines for little benefit. They felt that the guidelines should offer sets of default template designs, which could then be adjusted to meet particular demands.

- The guidelines made the group rationalise their media selection decisions. The novice designers were initially enthusiastic to animate all elements of the presentation, and to show colour photographs. However, when applying the guidelines it became clear that much of the motion information was non-complex, and could be more usefully communicated as still images (SR5), accompanied with language describing the actions. The method suggested also that in most cases simple line drawn images would be sufficient (IP1). The only video finally used were to locate the dip stick, which required the use of colour and texture cues in the crowded engine bay (IP2).

- Within media design guidelines were applied by the group. Considerable use of labels (IA5) , symbols (IA3) and highlighting (IA1) were made in the presentation. They also made use of zooms and pans to locate the dipstick (AP5).

- The guidelines made the group consider how the contact points should be designed between speech and visuals. Considerable effort was made in authoring the speech track. Rather than simply lift the original text from the maintenance manual and read it over their presentation, the designers instead redesigned much of it so that it tightly referenced what was shown in the presentation (CP1). This included using the same wording in labels and captions as in the speech track (VR9), with intonation for important details to in ensure that speech described what was currently being shown in their diagrams (SA4).

- The novice designers made effective use of the attention graph in deciding what should be in focus and when. Time was taken discussing which object should be in focus at which point in time, and deciding upon which presentation technique to use to move focus from one place to another. They utilised the attention graph to inform the design of labels and symbols to draw attention to particular objects from their diagrams. Reading time was allowed when labels were shown, generally by using the speech track to read out the label.

- The attention graph also promoted the use of reveals in the presentation. Whilst their initial decisions were to show everything at once, as they worked through the attentional guidelines and built their presentation graph, the issue of sequencing became important.

The novice designers went to considerable effort to show each label in turn, guiding attention around their presentation from one object to another (IA5).

- The group felt that the attention graph could be made even more useful if it provided more explicit advice on duration of media. They found it confusing to have produced the attention graph, but then to have to make further adjustments to it when it became clear that images would have to be shown for several seconds, and reading time left for text. Their first draft of attention graph showed on what was in focus, not for how long; this required re-drawing when the diagrams and labels were sequenced.

6.4 Adjusting the contact breaker points

6.4.1 Task Model

The contact pointers sequence described how the gap between the points could be adjusted using a feeler gauge. It required a sequence of actions to release the points, use the feeler gauge to set the correct gap, and re-tightening the points.

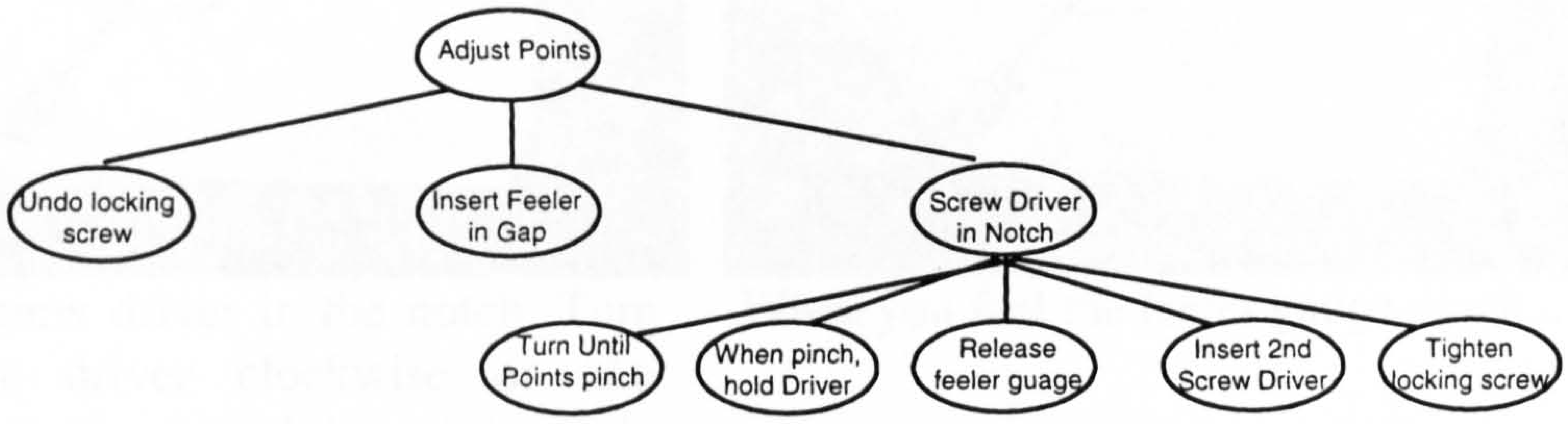
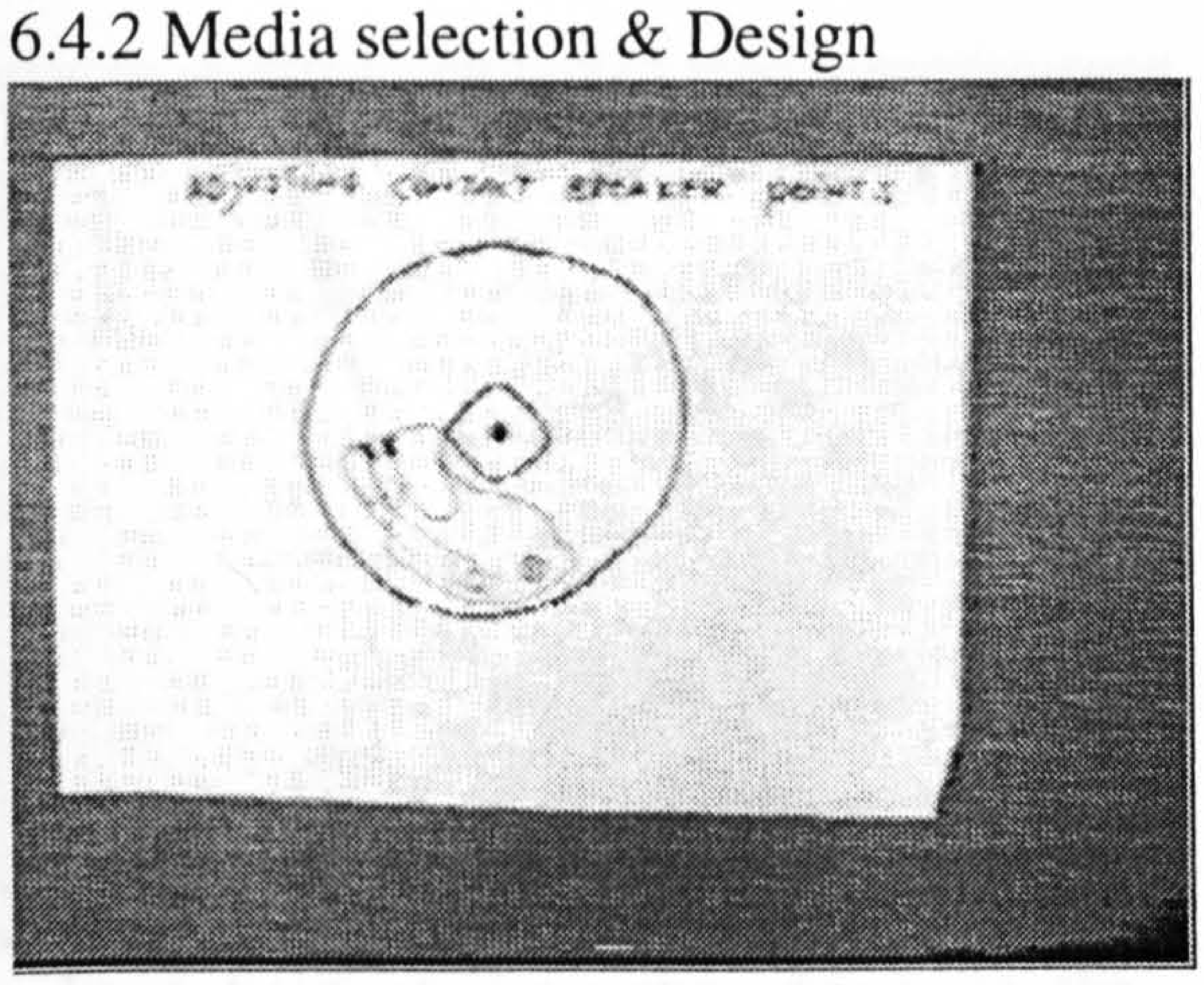
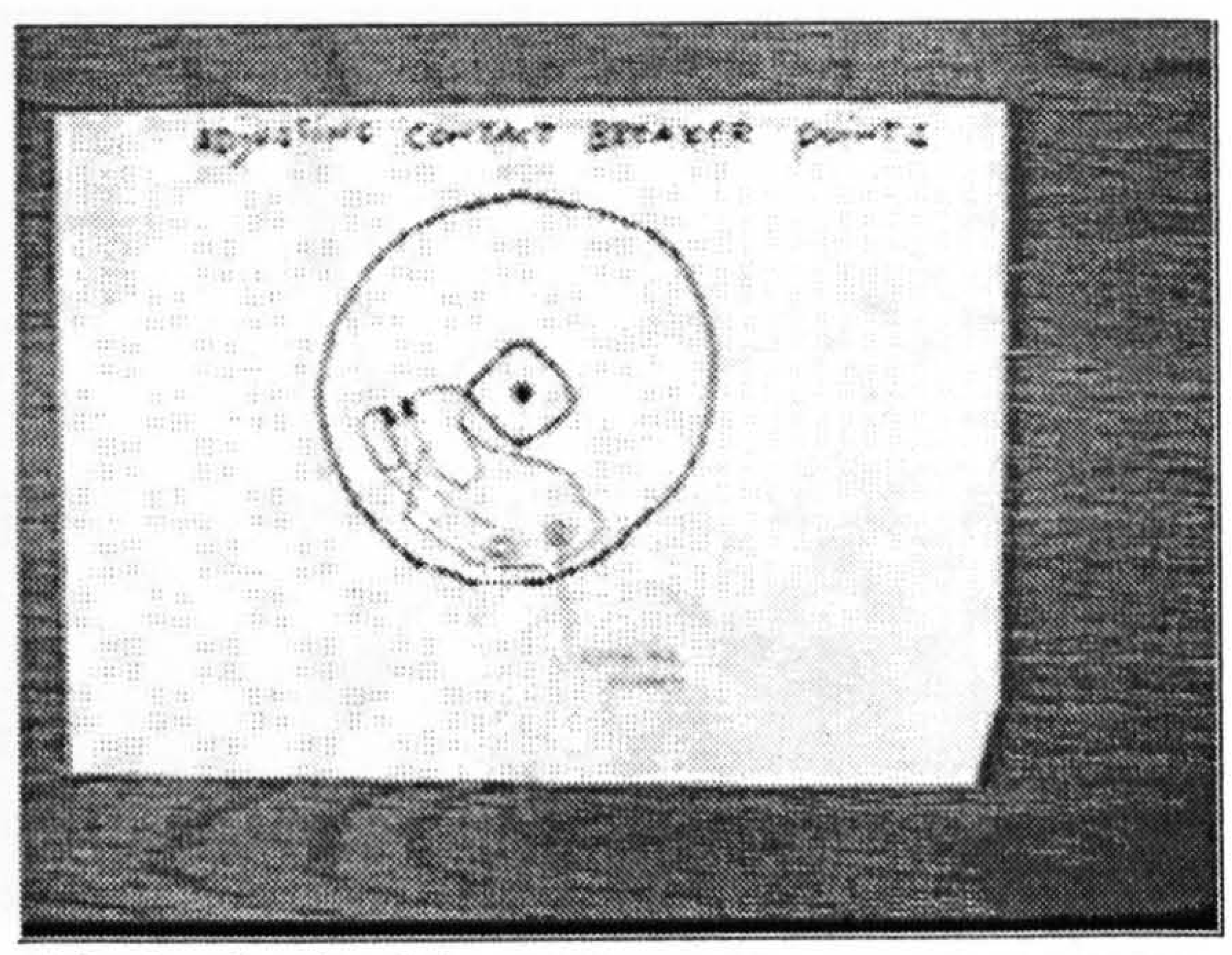


Figure 6.3.1 Task model for contact breaker task

6.4.2 Media selection & Design

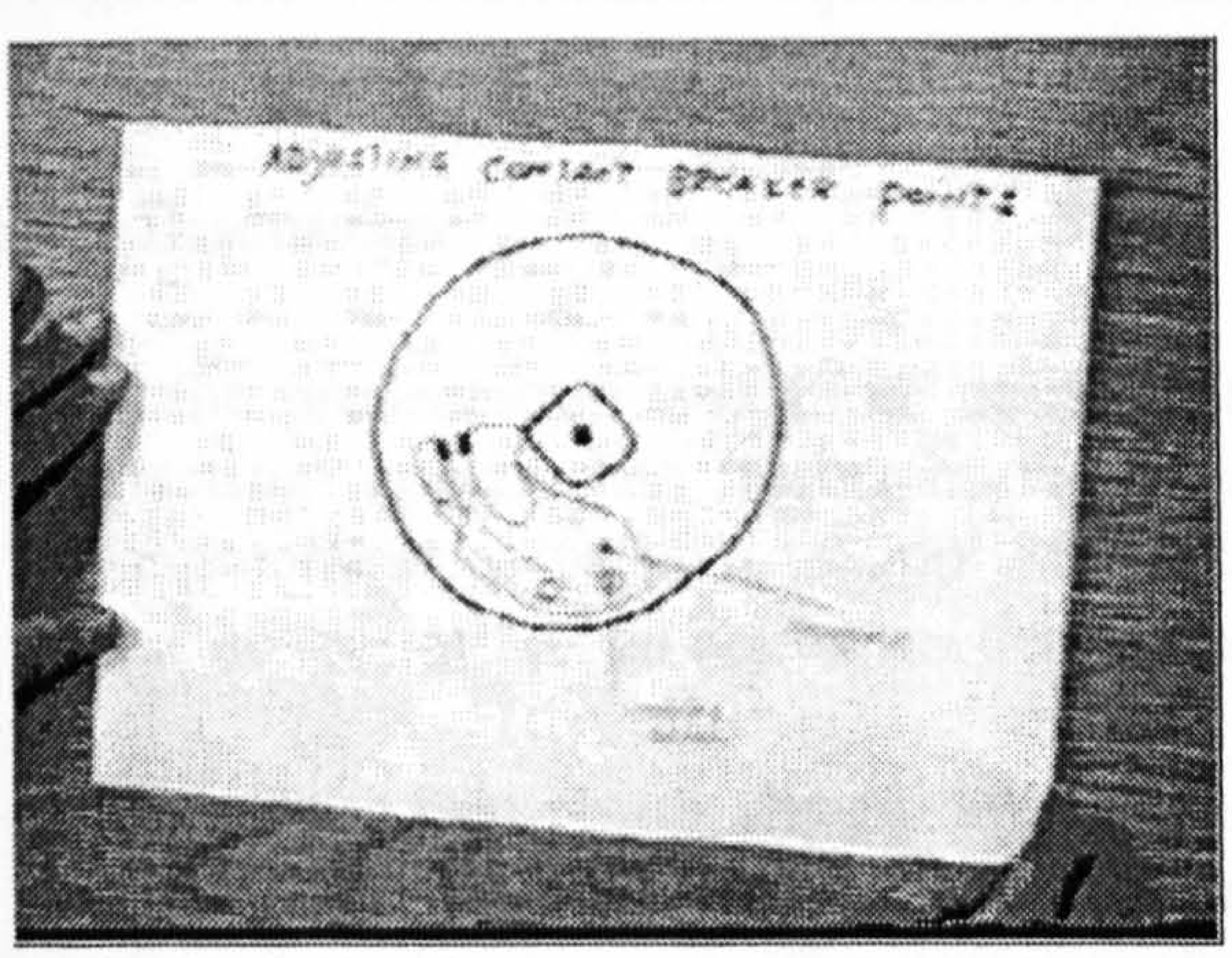


Adjusting contact breaker points.

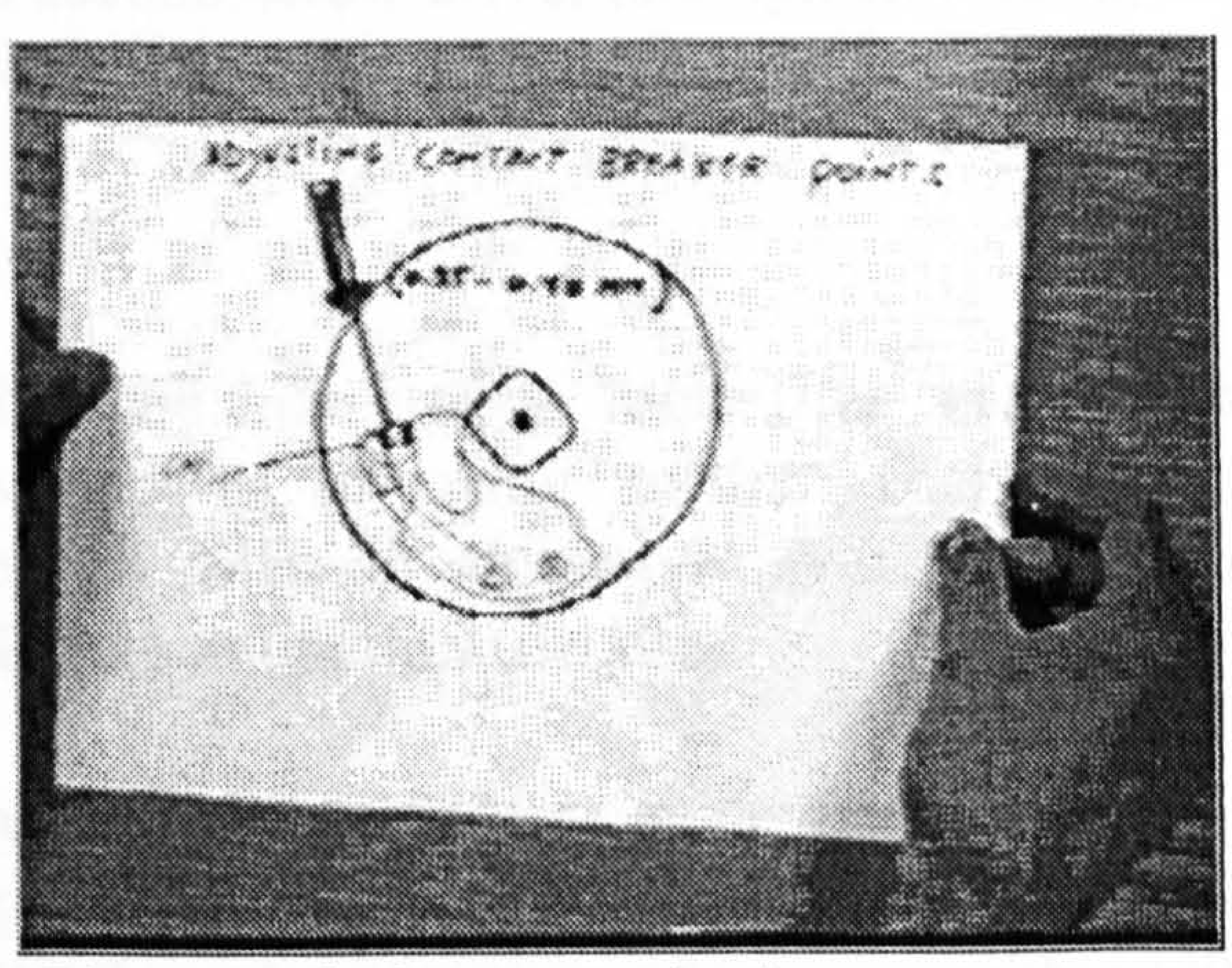


Undo the locking screw...

Figure 6.3.2 Presentation sequence for undo locking screw

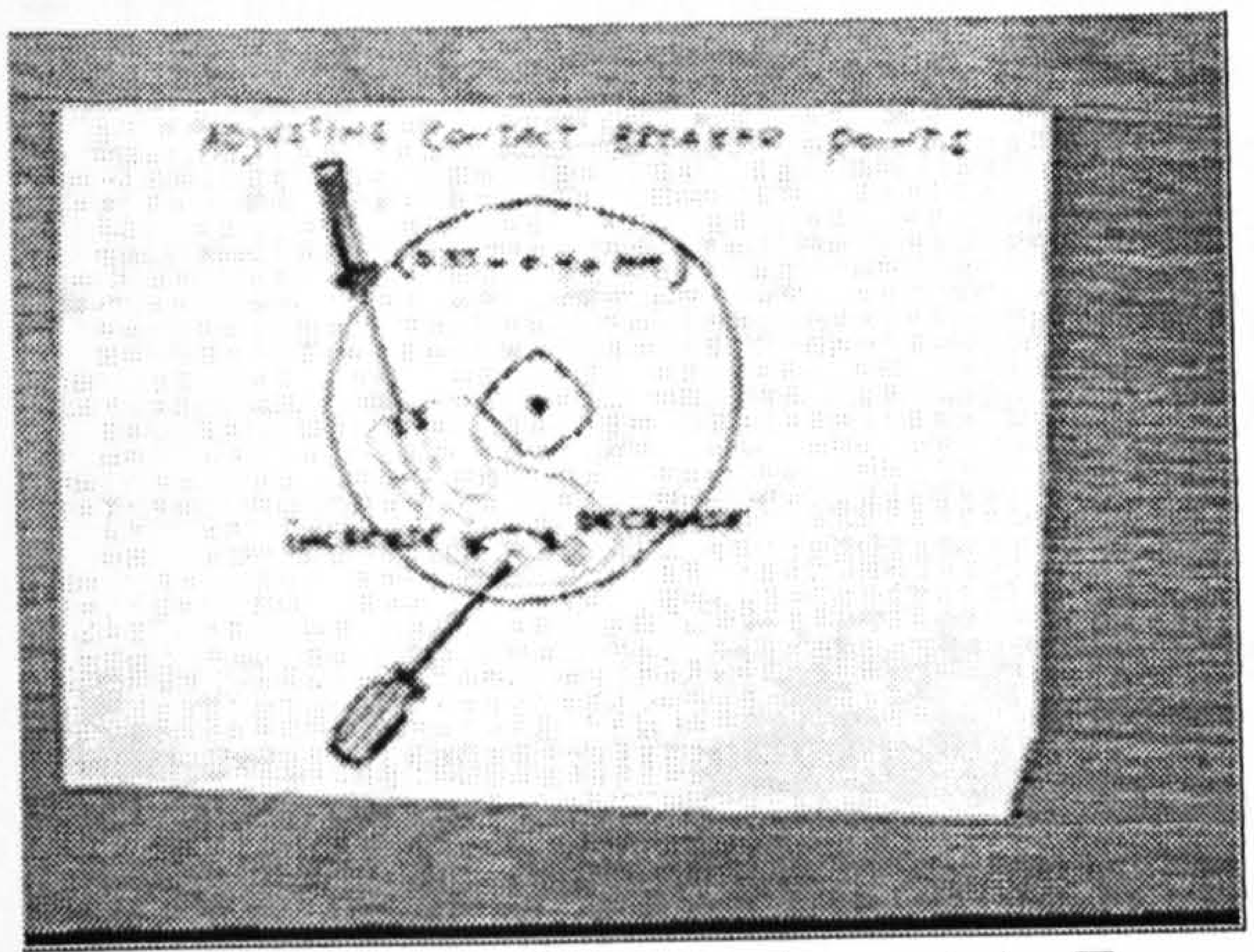


...in the base plate

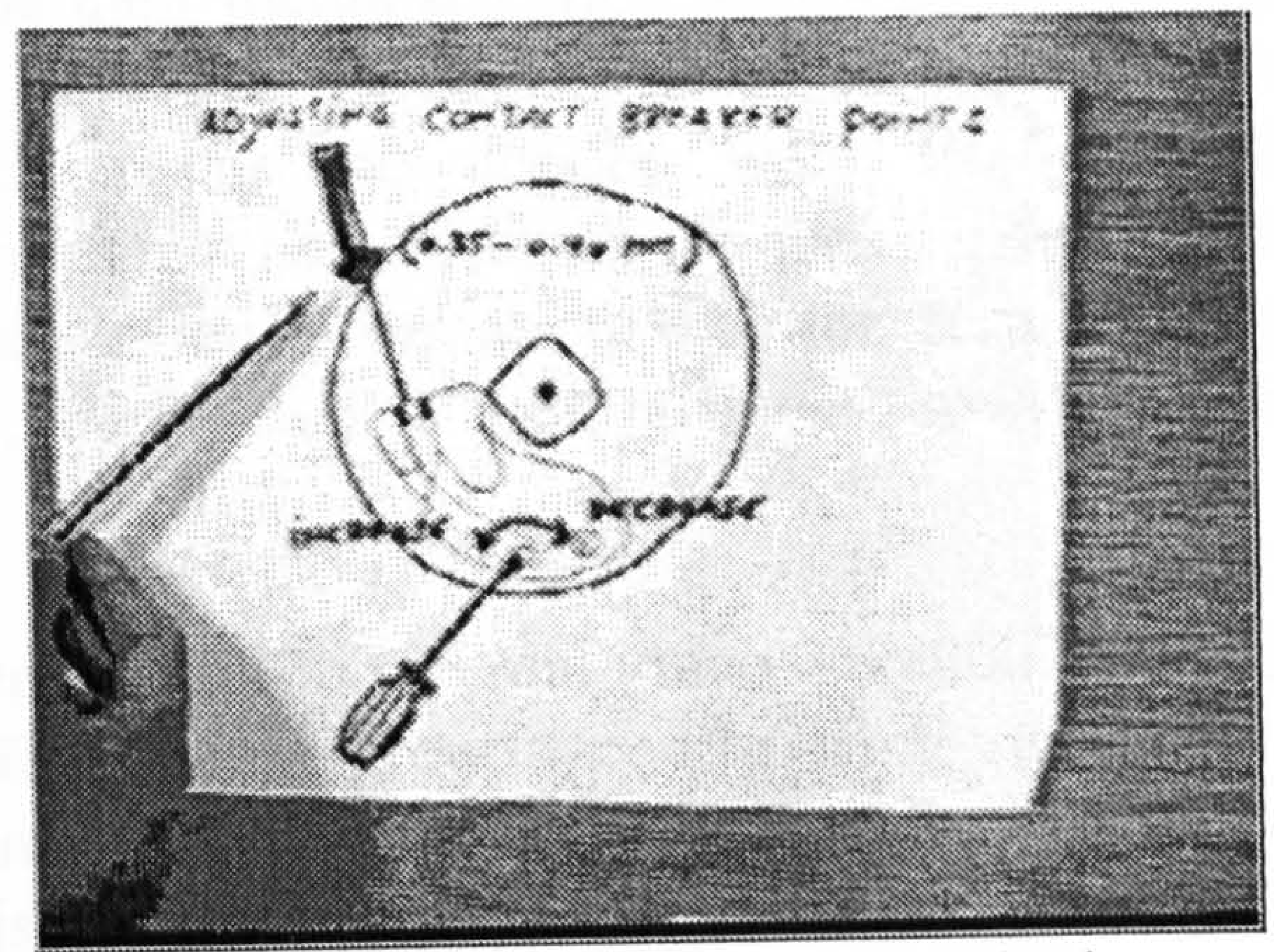


Place the feeler gauge in the gap.

Figure 6.3.3 Presentation sequence for insert feeler gauge

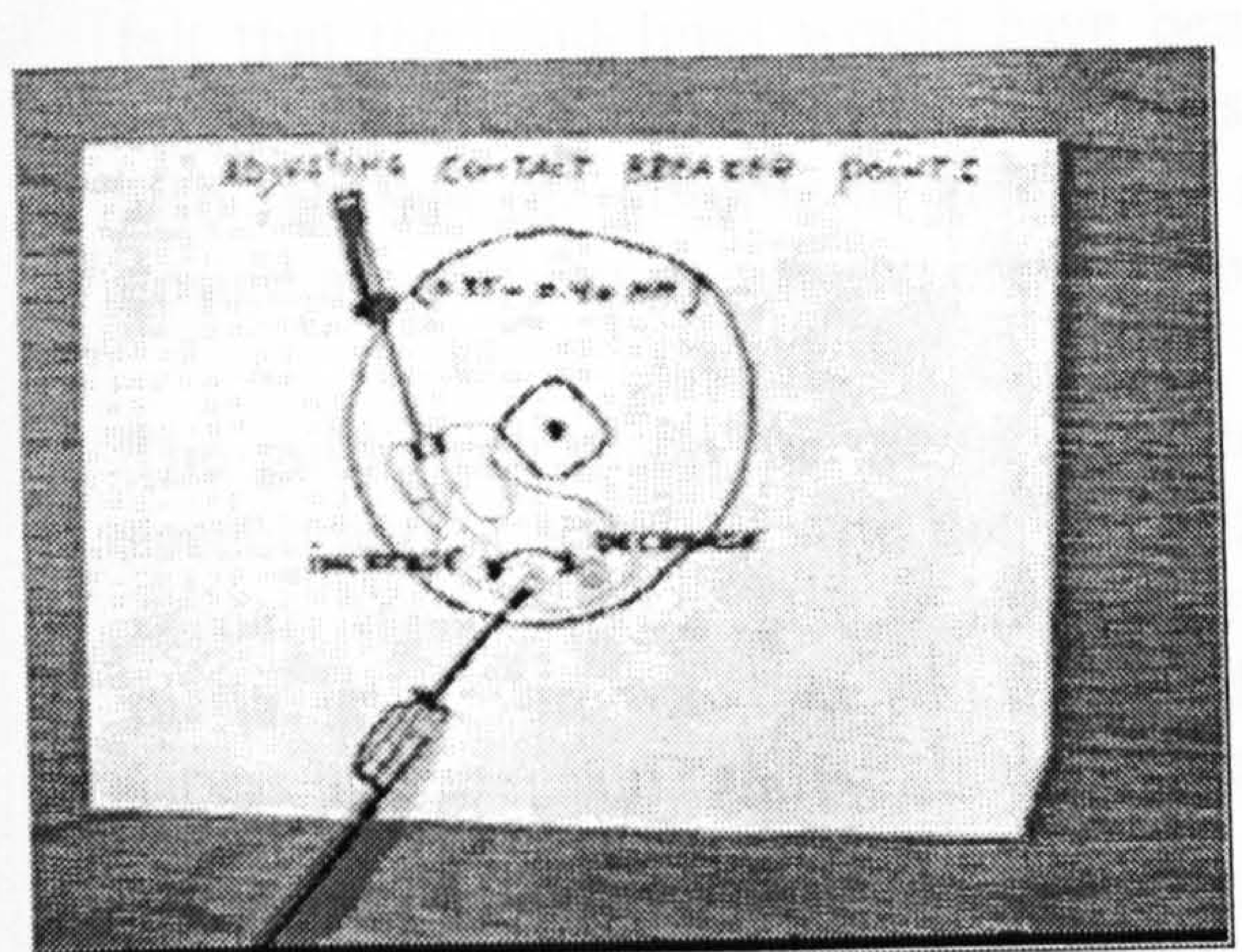


Place a screw driver in the notch. Turn the screw driver clockwise or anti clockwise until you feel the points pinch.

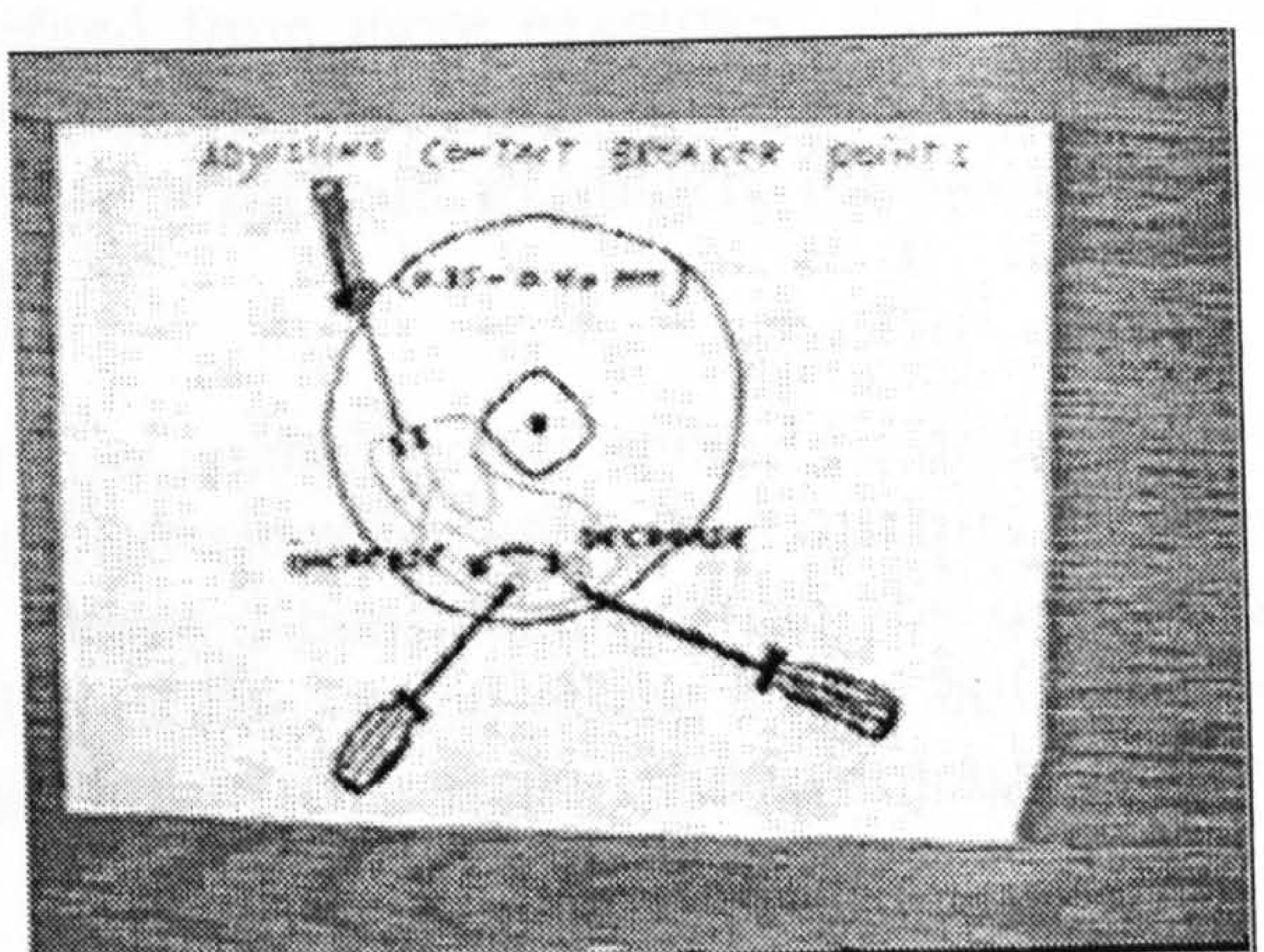


When you feel the feeler gauge pinch...

Figure 6.3.4 Presentation sequence for turn until point pinch step



...keep hold of the screw driver in the notch. Release the feeler gauge



Insert a second screw driver and tighten the locking screw

Figure 6.3.4 Presentation sequence for insert second screw driver and tighten locking screw steps

6.4.3 Discussion

- As the presentation was more complex than the first groups, more time was required in selecting appropriate information types and in choosing media. Their was some disagreement as to how to apply the selection rules for animated media, with confusion over how complex an action needed to be before animation was used.

- As with the first group, the guidelines were used successfully to guide media selection. The original presentation used a single all purpose illustration of the assembly, which was then referred to part by part. The designers instead used a sequence of images, with additional text and speech materials.

- The guidelines made the group rationalise their media selection decisions. Animation was used where the task information needs required it (SR5). Where the sequence required complex operational information, such as if two actions were to be performed simultaneously, e.g. tightening the locking screw and releasing the feeler gauge, animation was used (SR5a). However, much of the motion information was non-complex, and was communicated as still images (SR5b), accompanied with arrows showing path, and language describing the actions.

- As with the first group, the designers decided not to use real video or colour photographs. In this case, this reflected pragmatic difficulties the group faced in using video equipment within the time allocated for the study. Instead they produced a presentation entirely as line drawn images, which the method advised would aid identification and location of the objects.

- Within media design guidelines were applied by the group. Considerable use of labels (IA5), symbols (IA3) and highlighting (IA1) were made in the presentation. The second group was more critical of the within media design guidelines than the first group. They felt that the guidelines would have benefited from more examples, and from additional scoping to make clear in which circumstances particular guidelines should be used. They also felt that the volume of guidelines made it difficult to quickly get advice on how to design a particular part of the presentation.

- The guidelines made the group consider how contact points should be designed between speech and visuals. As with the first group, considerable effort was made in authoring the speech track (CP1). The wording was changed considerably from the original source materials. Effort was taken in the wording of the speech material so that it matched the labels and diagrams shown. Intonation and stress were again used to emphasise important objects.

- The second group ran out of time in producing their presentation, so they used the attention graph less than the first group. They did however apply the attentional guidelines. Effort was taken in highlighting objects in the image, using a pointer to show which should be focused on. Reveals were also used for labels to shift attention from one part of the presentation to another.

6.5 Conclusions

Whilst it is difficult to gain access to designers, it is important to take into account their feelings towards design methods. If they do not use the method then its value is lost. It is important to ensure that studies are conducted upon the community who are likely to use the guidelines.

The study had several criteria for success :

i) Novices should make use of each of the method steps to produce a satisfactory multimedia design

The study showed that the method could be used by novice designers, and that it helped them consider the choice of media to use, how to design within the media, how to construct contact points between speech and visuals, and how to design for attention.

Both groups could also rationalise and justify more effectively why they had made particular design decisions. When reviewing each others presentations, they proved to be able to critique and question at a much more detailed level than prior to their experience of the method. In particular, they no longer just stated that a presentation was poor, but were able to pick out what design changes they would have made to fix the problem.

The success of the method could be measured particularly in the first group, whose initial enthusiasms for animation, video and colour photography were redirected by using the guidelines. Whilst they shot around 5 minutes of video, in the end they choose to use simple line drawn diagrams and labels.

ii) The method could be applied by groups who were not connected with its development, and who had little background knowledge in multimedia design.

The criticisms the groups levelled at the method were useful, and showed how the method could be developed. The first group noted that the method asked them to keep asking the same questions over and over. They wanted some of this repetition removed with more general design templates, which could then be modified. These would encapsulate general presentation goals, such as to inform the user of how to locate an object, or perform an action.

Both groups disliked having to consult a lengthy document in order to get design advice. They were critical of the time it took to find the relevant material, and the difficulty in pin-pointing exactly which guidelines they needed to apply. They suggested the need for tool support to aid in using the guidelines and to partially automate validation of their design.

On several occasions they were also sceptical as to whether they really believed the guidelines advice, requesting clarification from the facilitator of why a particular design guideline existed. They felt they needed examples of good and bad design to illustrate the guidelines and that the guidelines be formulated to scope their advice more fully.

The first group also criticised using attention graphs without any reference to the duration of the media resources that were being sequenced. They complained that they had spent considerable effort constructing the representation, only to find that in fact it still meant that they would have to re-draw it as they produced their speech and diagrams. They suggested that the attention graph was useful once the media resources had been produced, not before.

iii) To understand the response of experienced designers to the method

The response of the professional designers to the exercise was far more critical than of the novice designers. They took great effort to point out that the guidelines were in places too specific, and in others too general to be of use. They felt that his approach to design was to 'just do it', and that guidelines would reduce creativity. When pressed to clarify why they felt certain guidelines inappropriate, they suggested that they needed greater background knowledge of the guidelines, and the studies on which they were based.

The professional designers were also highly critical that the guidelines ignored the aesthetic appeal of a presentation. They felt that by applying the guidelines laboriously, the presentations produced would be uninspiring and dull. They suggested that the creative knowledge of design he had gained allowed him to work more freely and innovate. They concluded that 'rules are meant to be broken'.

Chapter 7

Author tool Design and Implementation

7.1 Introduction

The aim of this chapter is to outline the design and implementation of the Author tool. The tool was designed to support the design method and to test how user driven support could be given in an authoring environment. This work is reported in [Faraday & Sutcliffe 97a], [Faraday & Sutcliffe 98].

The tooling issue for multimedia design support can be seen as having two separate components :

i) Authoring support for task, media, timeline manipulation, and presentation output,
ii) Guideline support for question / answer design advice, critiquing

The first of these issues, authoring support, is partially what conventional, commercial authoring tools supply. Many tools such as Director, Authorware etc. allow media to be input, some allow a presentation structure to be built up, many require a time line to be constructed, and finally use these facilities to provide some form of presentation play back.

The second issue concerns how guidelines can be used within this process. These will support the selection of particular media for a task step, the design within the media using presentation effects, and the attentional design of the presentation sequence over time. Few tools at present provide guidelines support. Some attempts have been made in AI work (see chapter 1) to allow presentation planning without intervention from the user, but none allow user authoring to take place with design support.

The need to combine the provision of guidelines with tool support has been highlighted in two studies of the effectiveness of paper based design guidelines. [Tetzlaff & Schwartz 91] provide a summary of problems with the use of a design guideline manual, and conclude that 'to assure conformity, guidelines should be developed in conjunction with the supporting environments and toolkits'.

A similar solution is also proposed by [Thovtrup & Nielsen 91], who performed a study in which designers were given an example interface to diagnose for problems using a usability manual. They conclude 'to increase the usability of user interface standards, we recommend having development tools that support implementation of interfaces that follow the standard'. These suggestions point to the need to also investigate how guidelines can be placed with the tool itself.

These results also agree with the problems found in chapter 6. In several cases the designers complained about having to consult lengthy method documents, and in the difficulties in using the paper based representation. A design advisor would thus seem to be of value as a way of the delivering the guidelines.

7.2 Existing tools

Current authoring tools can be divided into two distinct paradigms. The first contains many authoring systems which allow the user to design an interactive presentation. These use a variety of metaphors for authoring, but provide no support for design advice or critiquing.

The second contains tools which provide guideline support, but at present none support multimedia authoring. Instead these tools are mainly aimed at traditional UI design. However, as the issues that these tools address are similar to guidelines support for multimedia, they are also of relevant.

7.2.1 Multimedia Authoring tools

Several different types of authoring metaphor have been proposed to aid in the design of interactive multimedia presentations. The following list is far from exhaustive, but does provide an overview of the current approaches in both commercial and research systems :

i) Structure based tools

In a structure based tool, the authoring system gives explicit support for the representation and manipulation of the content structure of the presentation. This allows the presentation content to be worked upon explicitly. Simple time relations can also be automatically derived from the structure of the content.

An example of a structure based tool is CMIFed [Hardeman et al 93]. It uses a hierarchy view to allow author to control structure of presentation, see figure 7.2.1. Timing information is given in a channel view; each media item is assigned a channel. It shows the timing relations given in the hierarchy view. Synchronisation is achieved by specifying parts of the hierarchy as parallel or sequential.

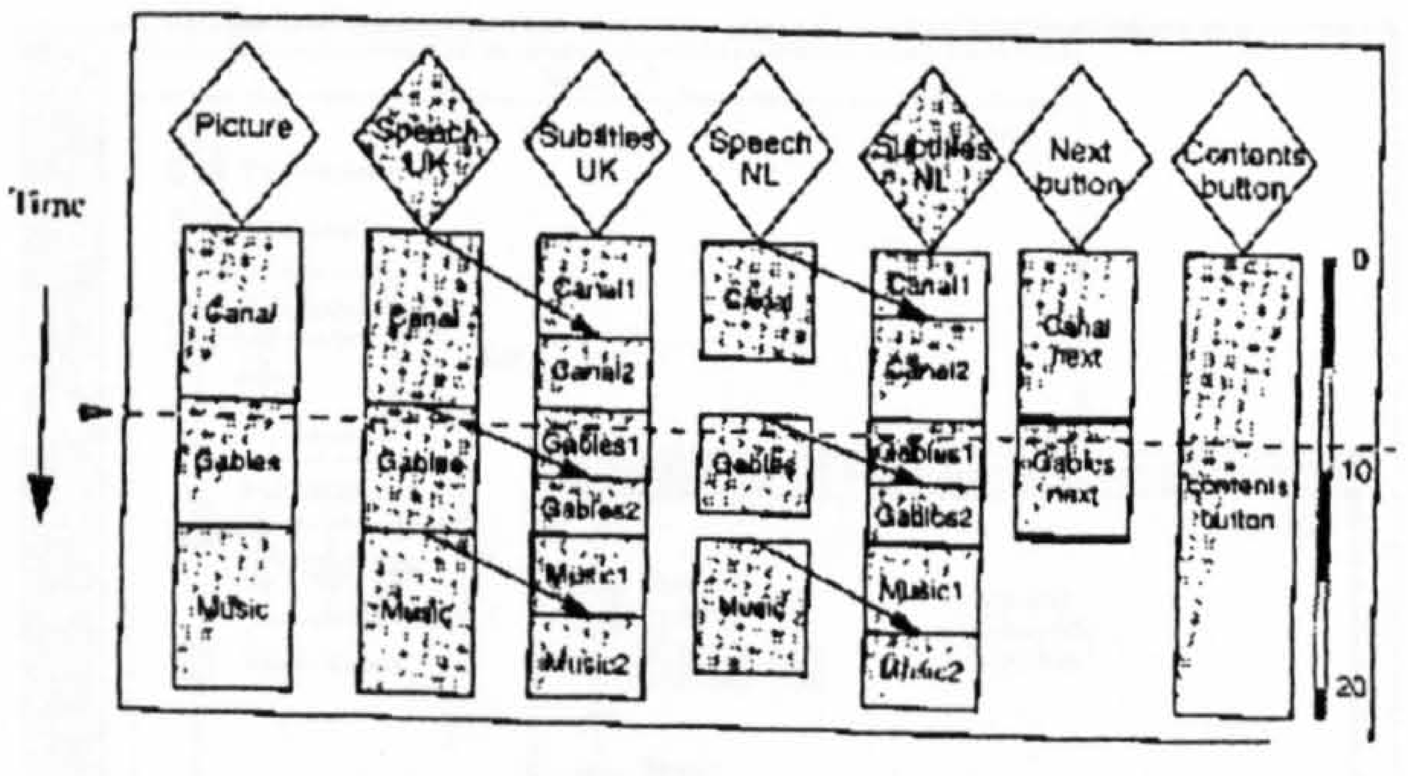


Figure 7.2.1 CMIFed Hierarchy View [Hardeman et al 93].

Structure based tools suffer from one significant problem. They can be difficult to use for complex time relationships since they do not show time relations, and synchronisation can be difficult, particularly of time varying media or for overlapping media in time

ii) Timeline based

Timeline based tools show the constituent media items placed over time. This may be on different tracks, to allow simultaneous presentation, or allow overlapping of media items. The time lines allow complex synchronisation of time varying media and overlapping. An example timeline tool is Director [Macromedia 96]. Media are placed on a timeline or 'score', shown in figure 7.2.2. The timeline has a number of tracks, allowing parallel and sequenced presentation.

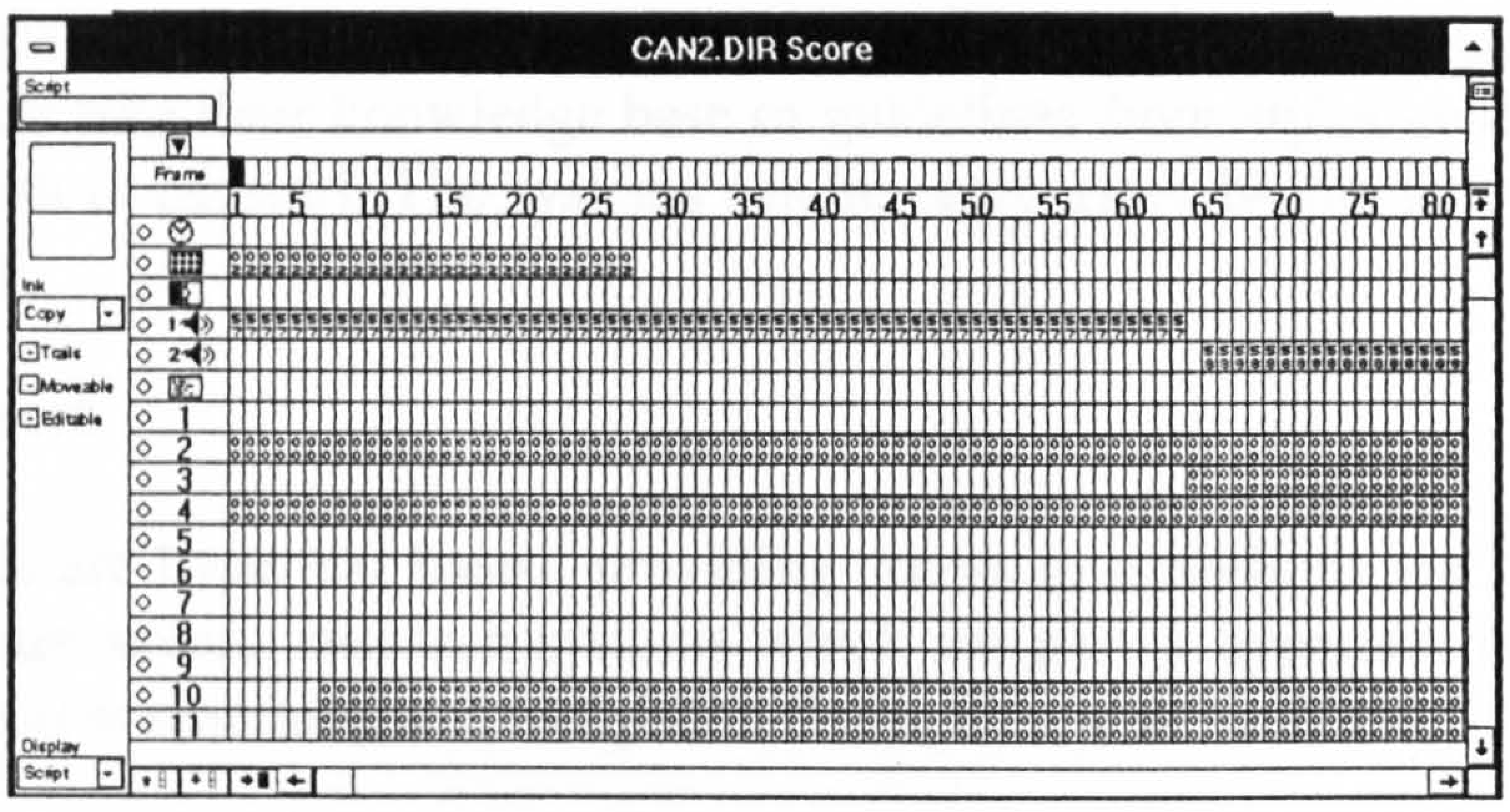


Figure 7.2.2 Macromedia Director Score view [Macromedia 96].

Timeline systems have two problems. First, the presentation is treated as a single mass, can rapidly become cumbersome. Second, no representation of the content itself is supported e.g. as a task structure.

iii) Flowchart based

A flowchart based tool gives the author a visual representation of the commands describing a presentation, and how they interact together. This allows content structure to be worked upon explicitly. An example flowchart based tool is [Authorware 96], see figure 7.2.3. Icons are selected from a palette an incorporated into a flowchart.

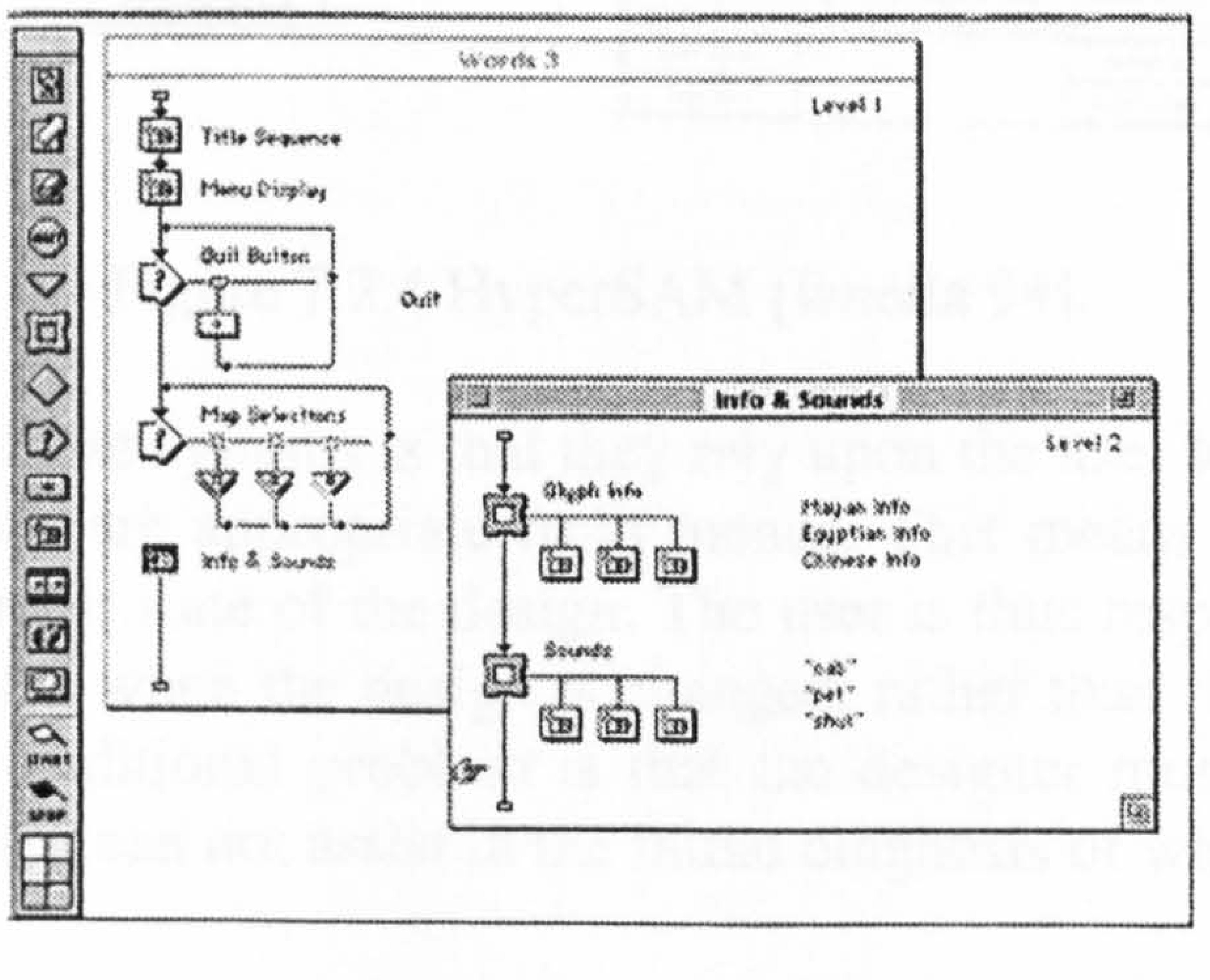


Figure 7.2.3 Authorware Flow Chart [Authorware 96].

A disadvantage with flowchart systems is that timing and synchronisation can be complex, particularly of time varying media or for overlapping media in time. Thus, in the case of [Authorware 96] there is no way of getting a timeline overview of which media will be shown when.

7.2.2 Design Advisor Tools

As noted, the majority of design advisor tools address more conventional issues in UI design. The tools generally take their knowledge base of guidelines from styles guides, such as those provided by [Smith & Mosier]. The systems can be categorised by the amount of intelligent support they offer.

i) Hypertext advisors

The simplest tools are hypertext based, providing access to guidelines via browsers, such as MS Help. Their are several research systems which are purely hypertext based. These take style guides and allow searching and navigation through their content.

Examples of these systems include SIERRA [Vandervonckt & Bodart 93] and HyperSam [Ianella 94], shown in figure 7.2.4. Both provide a hierarchical organisation of design advice, allowing the designer to search particular topics of interest e.g. HyperSAM supports four methods of interaction with SAM guidelines report : browsing, gathering, annotating and searching. SIERRA is similar, in that it provides a hypertext set of links between various subsets of guidelines.

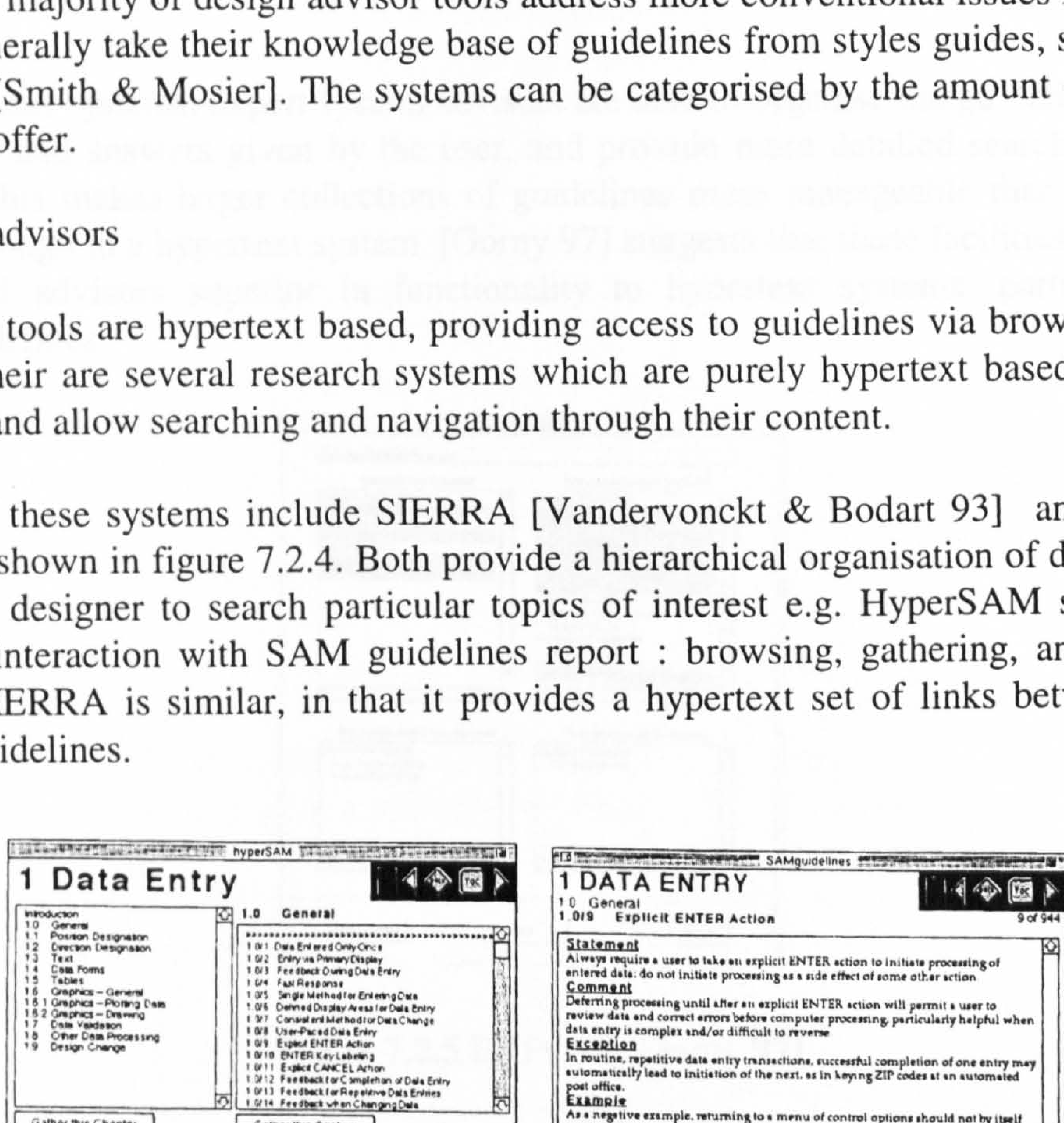


Figure 7.2.4 HyperSAM [Ianella 94].

The main problem with these systems is that they rely upon the user to consult with them and to select which guidelines are appropriate from menus. This means that the advisor has no way of accessing the current state of the design. The user is thus responsible for making sure the advisor is kept updated when the design is changed, rather than the advisor being able to monitor the design. An additional problem is that the designer must be aware of potential problems, since the advisor can not assist in the initial diagnosis of which guidelines to apply.

If a large number of guidelines are held or if many guideline are related together, hypertext systems can also be time consuming to search through. This is because it is left to the user to decide which guidelines they have already viewed, and to navigate through the advice.

ii) Experts system Advisors

These tools are more complex, making use of inference engines to provide support in accessing the guidelines. They allow diagnostic support, such as asking what-if style questions.

Unlike hypertext systems, expert system advisors are able to organise the guidelines based on the questions and answers given by the user, and provide more detailed search for relevant guidelines. This makes larger collections of guidelines more manageable than if they were navigated through in a hypertext system. [Gorny 97] suggests that these facilities make expert system based advisors superior in functionality to hypertext systems, particularly with complex guidelines.

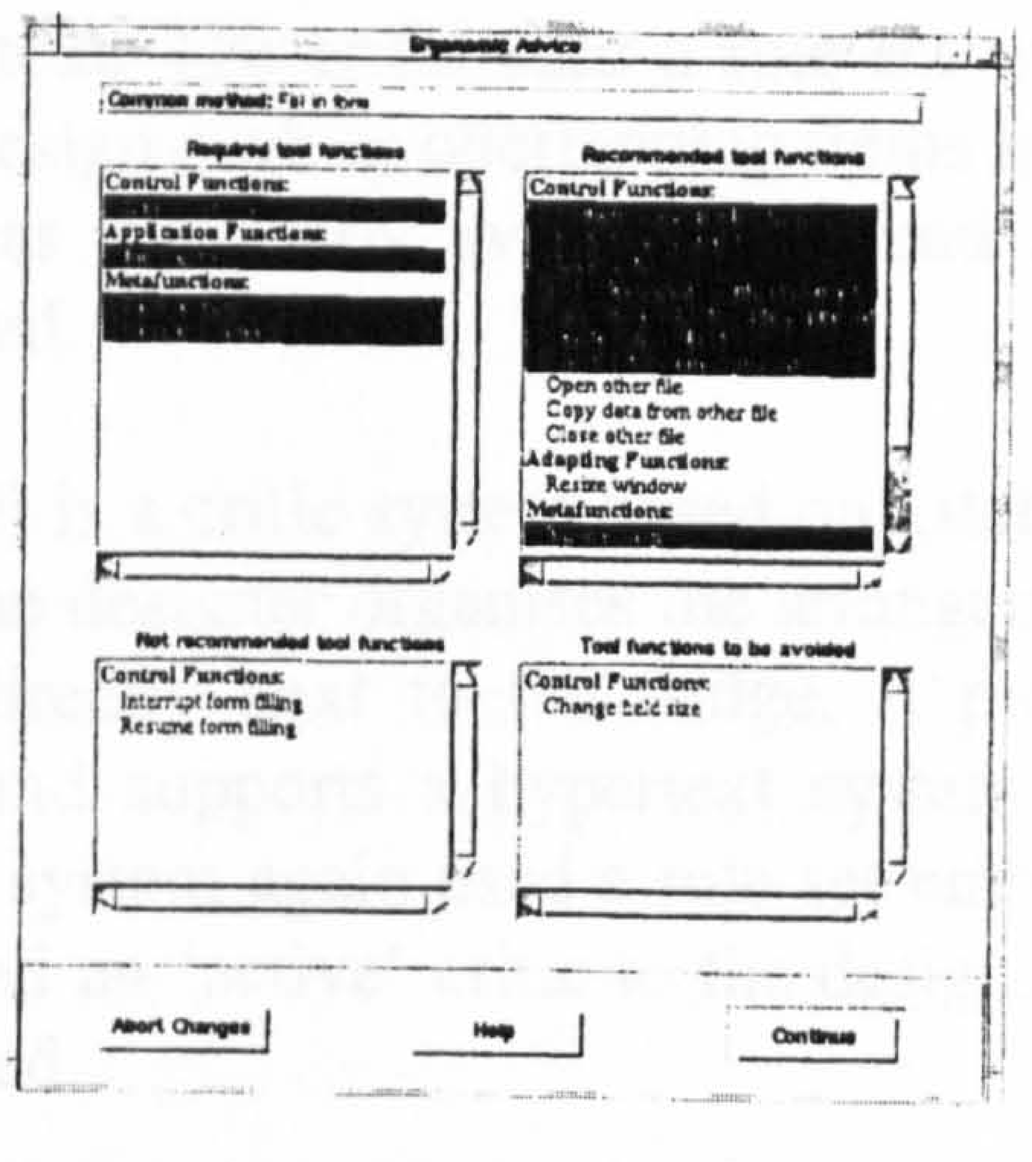


Figure 7.2.5 EXPOSE [Gorny 97]

EXPOSE [Gorny 97] is an expert system based on rules for UI design. EXPOSE provides more support than IDA, since the user builds a model of the task within the system e.g. for an order system, the user would set up an order form object as part of a form fill task step, see figure 7.2.5. The user then selects which type of interaction is required, and the tools suggests a suitable set of UI methods such as Fill_in_data, Close_form etc. Having set up the elements of the interface, more specific advice can then be gained on ergonomics and UI design. The system also provides a hypertext advisor to present more detailed advice.

The only current system which exploits expert based design specifically for multimedia advice is eMMa (environment for multimedia Authoring) [Nakakoji et al 97]. The eMMa system provides advice for the design of images e.g. which colours to use and in which proportions. The system allows the designer to input a viewer profile, and will advise on the use of colour and images based on it e.g. for 'audience : customer', 'style : festive' and 'atmosphere : cheerful' the system would suggest using a yellow colour. These colours can then be applied to the image.

A significant problem with expert system based tools is that the user must regularly consult them and update their representation of the problem state. This can be costly, particularly is the user does not constanly update the expert system as the design evolves leading to large gaps betwen the current state of the design and the expert systems state.
iii) Critic based

These tools are generally built upon expert system tools, but are implemented as part of the UI design environment. This enables them to synchronise automatically with the state of the design, and to inform the designer of any violations of guidelines without the designer being required to act. This facility allows a critic system to deliver advice appropriate for the designer's current problem or area of focus.

Critic systems doe not fix the problems they find, but point them out to the designer. They may also contain hypertext sub-systems to provide specific details of the design problems.

The first critic based system built was FRAMER [Fischer & Lemke 90]. It provided a critic facility within UI design system. The critic used a rule set within an expert system to spot problems in the state of the design such as overlapping items in the display, or a lack of a title for a window. The critic was constantly available, termed an 'active' critic, and spotted problems as the design evolved.

JANUS-CRACK [Fischer 90] is a critic system based on kitchen design. It is able to apply a set of guidelines in-situ, as the designer organises the arrangement of design components e.g. the cooker should not be directly next to the fridge. It provides real-time feed back of potential design problems, and supports a hypertext system to give more detailed design advice, see figure 7.2.6. The system again used a rule set embedded within an expert system and as with FRAMER applied an 'active' critic to the design, informing the designer of any problems as the design evolved.

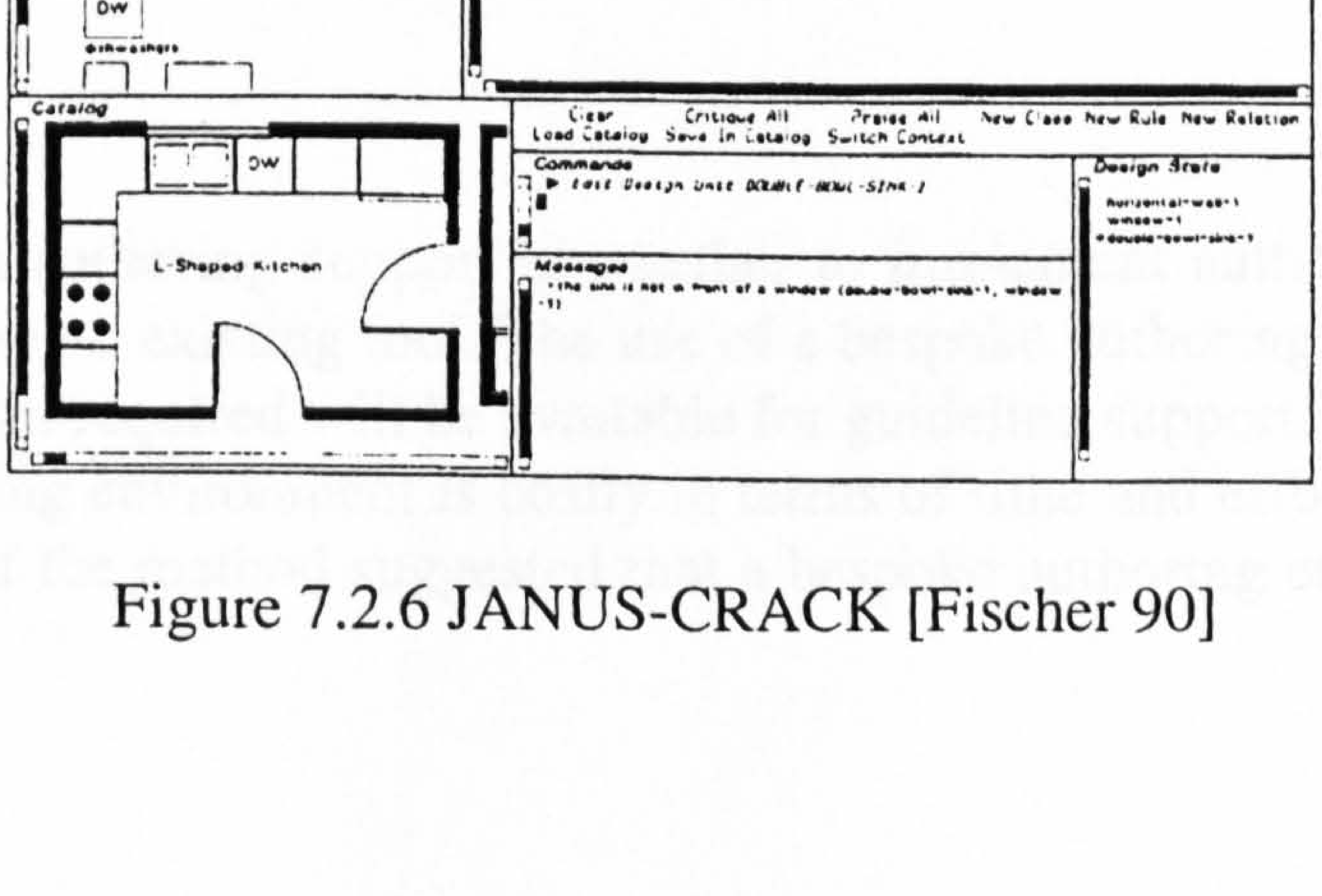


Figure 7.2.6 JANUS-CRACK [Fischer 90]

[Lowgren & Norquist 92] note several problems with critiquing systems. First they suggest that a passive critic may lead to problems if a large number of design errors are made before the critic is consulted. They conclude that there is a danger that the critic may overwhelm the user with irrelevant advice. They propose that the [Fischer] system of active critiquing is a partial solution to this, since it will help to spot problems earlier.

Second, [Lowgren & Norquist 92] point out that it is difficult to accurately formulate rules which are specific enough to diagnose problems usefully, but which are not so general as to constantly annoy the designer. In many areas of UI design, it can be difficult to accurately specify such guidelines.

7.2.3 Summary

The many different metaphors for multimedia authoring are less important in themselves than identifying a metaphor which will be useful for an expert system to interrogate. The following key issues are raised by current guideline tools :

i) The tool should combine a timeline view with a structure view of the content. As noted, the timeline is useful for accurate synchronisation, which will be required for attentional design; whereas the structure view would make it easier to understand how the task is being organised and aid in media selection and within media design.

ii) An expert system based tool would seem to be more useful than a pure hypertext system. If the rules are complex, then hypertext would rapidly become unwieldy. Also the expert system tool can more specifically target advice. A model of the task, media and presentation effects is essential. These must be built in to a tool that is accessible to the expert system advisor.

iii) The advisor should work in two ways : advice and critique. Current guideline tools make much out of the ability to both provide detailed advice on demand (usually via hypertext), and to also apply design rules as when the user requires a critique. The critiquing mode would seem to be particularly useful in freeing the designer from having to constantly query the the expert system for advice; instead design problems should be notified to the user as they occur. The critic should be active, alerting the designer to potential problems as the presentation evolves. The guidelines should be used as basis for producing critquing rules.

7.3 Building Authoring support

7.3.1 Introduction

The first key issue in authoring support is whether to implement authoring facilities within the tool itself, or to use an existing tool. The use of a bespoke authoring support environment means that the facilities required will be available for guideline support. The downside is that developing an authoring environment is costly in terms of time and effort. Several key issues in the requirements of the method suggested that a bespoke authoring environment would be useful :

i) In order to provide guideline support, the authoring environment had to be open and be able to be manipulated. At the time of implementation, no existing authoring tools met this need.

ii) The method requires a model of the user's task. Few commercial authoring environments allow the users task to be input explicitly into the authoring environment.

iii) The method requires a model of the media resources. No current tools allow the user to input explicitly the procedures, objects and actions and build the presentation out of them. Many tools have a concept of 'cast', but do not allow media resources to be grouped into particular information types, e.g. a text label, caption and part of a bitmap to be grouped together as a particular object, which is then used in an action.

iv) The tool needs to support a set of presentation effects such as highlight, label, zoom. Whilst many tools allow the user to perform an almost unlimited number of effects upon media resources, few provide explicit types of effects or structure these effects to allow the tool to aid in their selection. The tool needed to support certain types of effects, which could then be applied to particular media resources as the result of certain guidelines.

The second key issue is what functionality would be required to support the guidelines component of the tool. Here there is a need to be explicit about what the guidelines would need access to within the authoring environment, and how the guidelines shape the way the authoring environment is used.

i) The guidelines should be rule based. Declarative style rules make the guidelines more explicit, allowing for easier maintenance. This requires an expert system style interpreter.

ii) The guidelines must be able to access, search and update the representation formed from the authoring environment in both a passive mode, driven by the user, and in an active critique mode, spotting problems in the presentation.

iii) The guidelines must be able to present advice to the user, to show warnings, and to update the author tool

7.3.2 Architecture

The prototype tool was built in Visual C++ and AMZI prolog under Windows 95. It has two main components, shown in figure 7.3.1. The authoring component is used to build a task / media model. This represents the task which is to be presented, the media which are to be shown, and their sequencing in time. The design advisor component uses an expert system which is able to interrogate the task / media model formed by the authoring tool and apply a set of design rules based on the guidelines. These are used to produce the attention graph and provide design advice. The following sections address key issues in the design of the authoring tool and expert system .

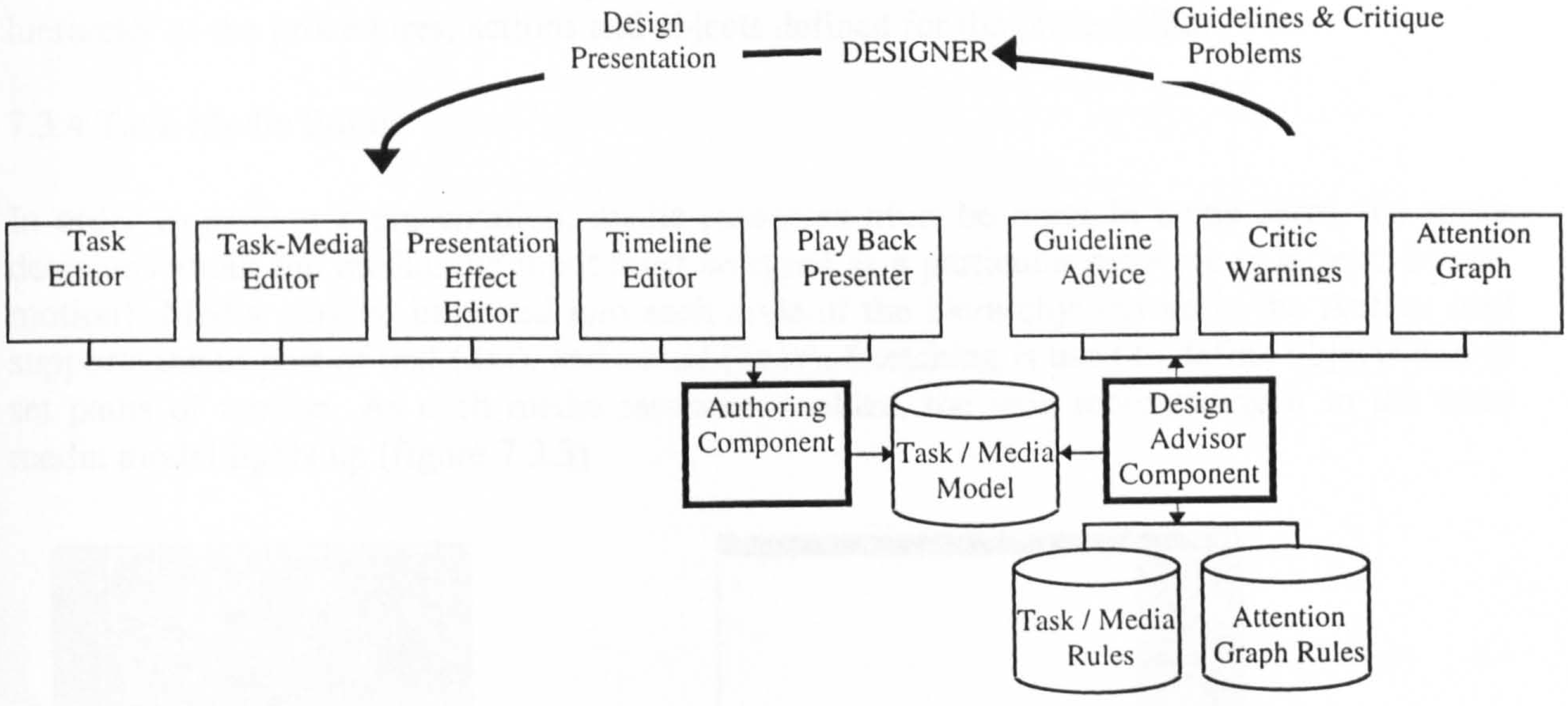


Figure 7.3.1 Overview of tool architecture

The review begins by outlining the design of the authoring component. This contains several elements which will be discussed in more detail below. First the design of the task and media models is discussed, along with the task editor and media editor which allow the user to enter the task representation and populate with media and presentation effects. In order to produce a presentation which can be re-played, the timeline editor is then used to set up time dependencies between the presentation effects.

7.3.3 Task Editor

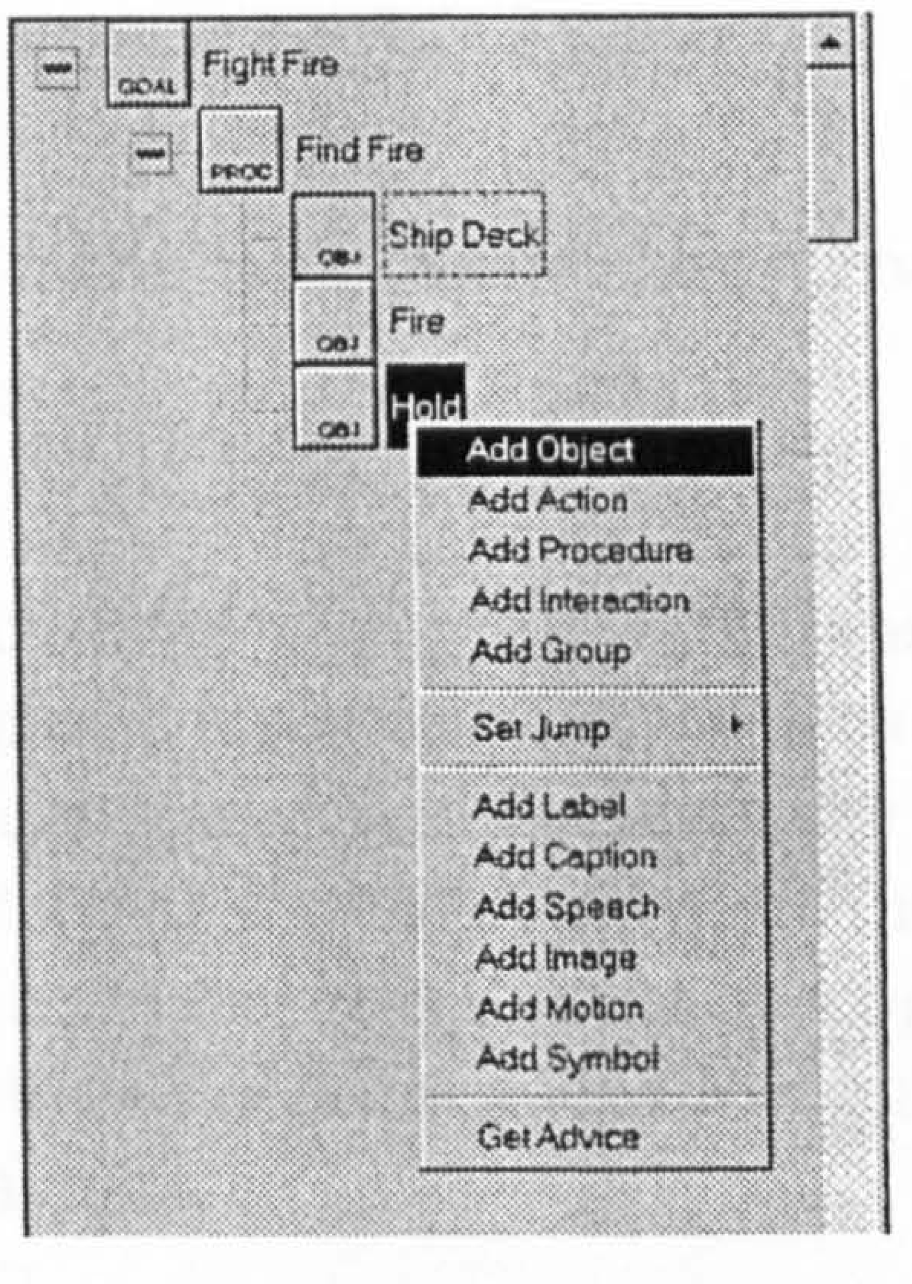


Figure 7.3.2 Partial task model for Shipboard emergency 'find fire' task sub-goal

The first input for the Author tool is the task specification. This takes the form of information types e.g. procedures, actions, roles, scenes and objects formed into a hierarchy to represent the task to be presented, shown in figure 7.3.2. The Author tool shows the task model as a hierarchy of the procedures, actions and objects defined for the presentation.

7.3.4 Task-Media Editor

In order to present a presentation, media resources must be input in some form. To make decisions about the media, the input must be typed as a particular resource (e.g. text, image, motion). Media may be imported into each node of the hierarchy; currently the Author tool supports the import of text (.txt), and sound (.wav). Sketching is used to define objects and to set paths of motion. As each media resource is added, the icon representing it in the task/ media model lights up (figure 7.3.3)

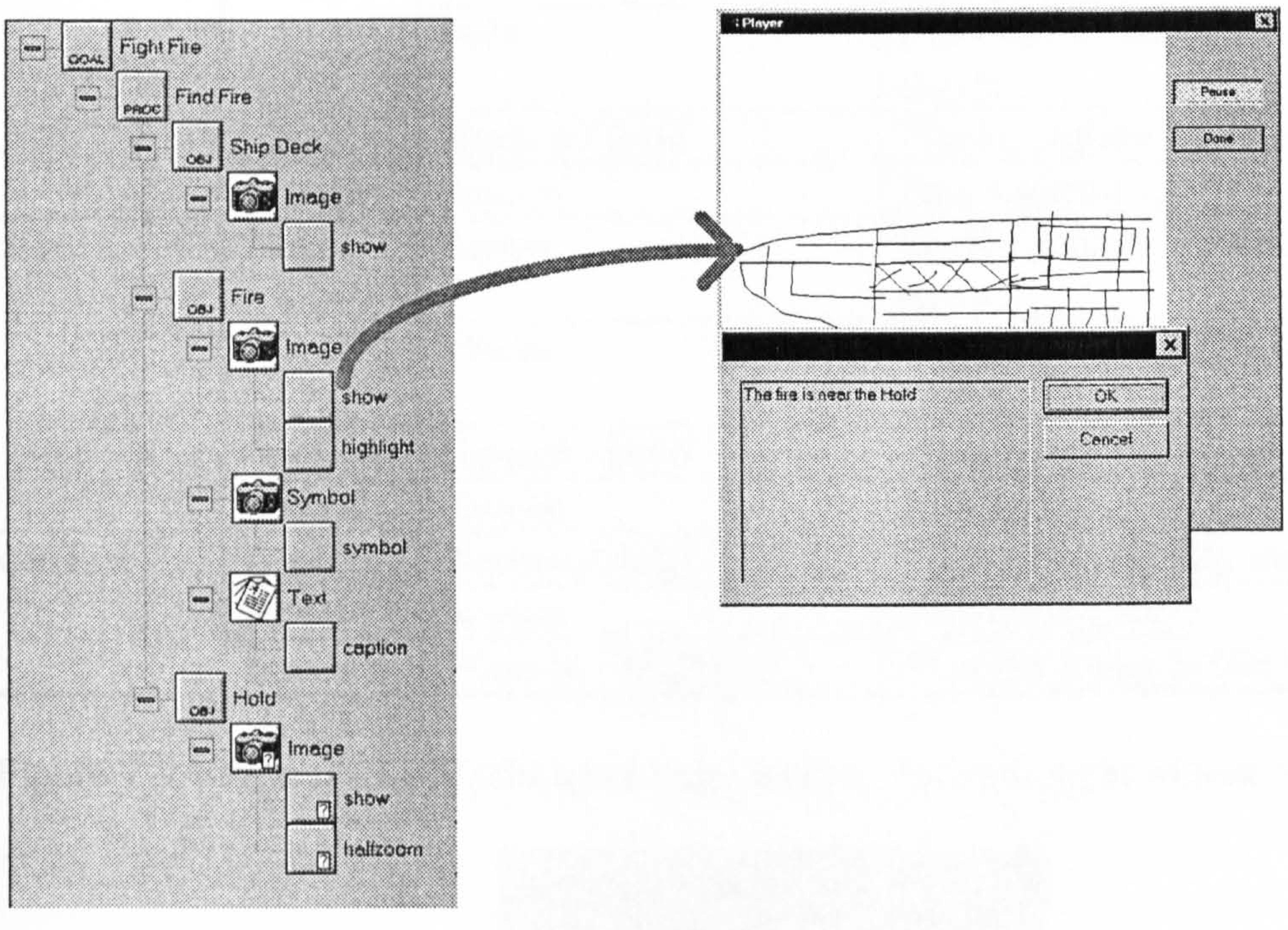


Figure 7.3.3 Task-Media model (left), Sketch based image of ships deck (right)

7.3.5 Presentation effect editor

Presentation effects can then be added to support within media design.These provide pre-defined effects upon particular media resource types, such as zoom-in, highlight, show path, wipe, pan. Presentation effects are placed into the hierarchy built in the task-media model. These are held as a list of effect types for each task-media node; thus an object node in the hierarchy could have a highlight and label effect. A list of supported effects is shown in figure 7.3.4. Figure 7.3.5 shows how the effects are added using the tool. Figure 7.3.6 shows an example of the effects during playback.

Task node type	Media Type	Presentation technique	Description
Object	Image	Zoom-in	Close up zoom upon an object in an image
		Zoom out	Zoom out to show the full image
		Highlight	Place a highlight over an object
		Symbol	Place symbol over an object
	Label	Normal / Bold	Place a label text over an object
		Bold	Place a bold label text over an object
	Caption	Normal / Bold	Place a caption into the scene
	Speech	Speech	Play speech back
Action	Image	Arrow	Show the path of motion as an arrow
		Freeze	Freeze the path of motion of an object
	Caption	Normal / Bold	Place a caption into the scene
	Speech	Speech	Play speech back
Procedure	Caption	Normal / Bold	Place a caption into the scene
	Speech	Speech	Play speech back
	Image	Wipe-in / Wipe-out	Turn the frame to black

Figure 7.3.4 Presentation Techniques supported for different types of task node and media

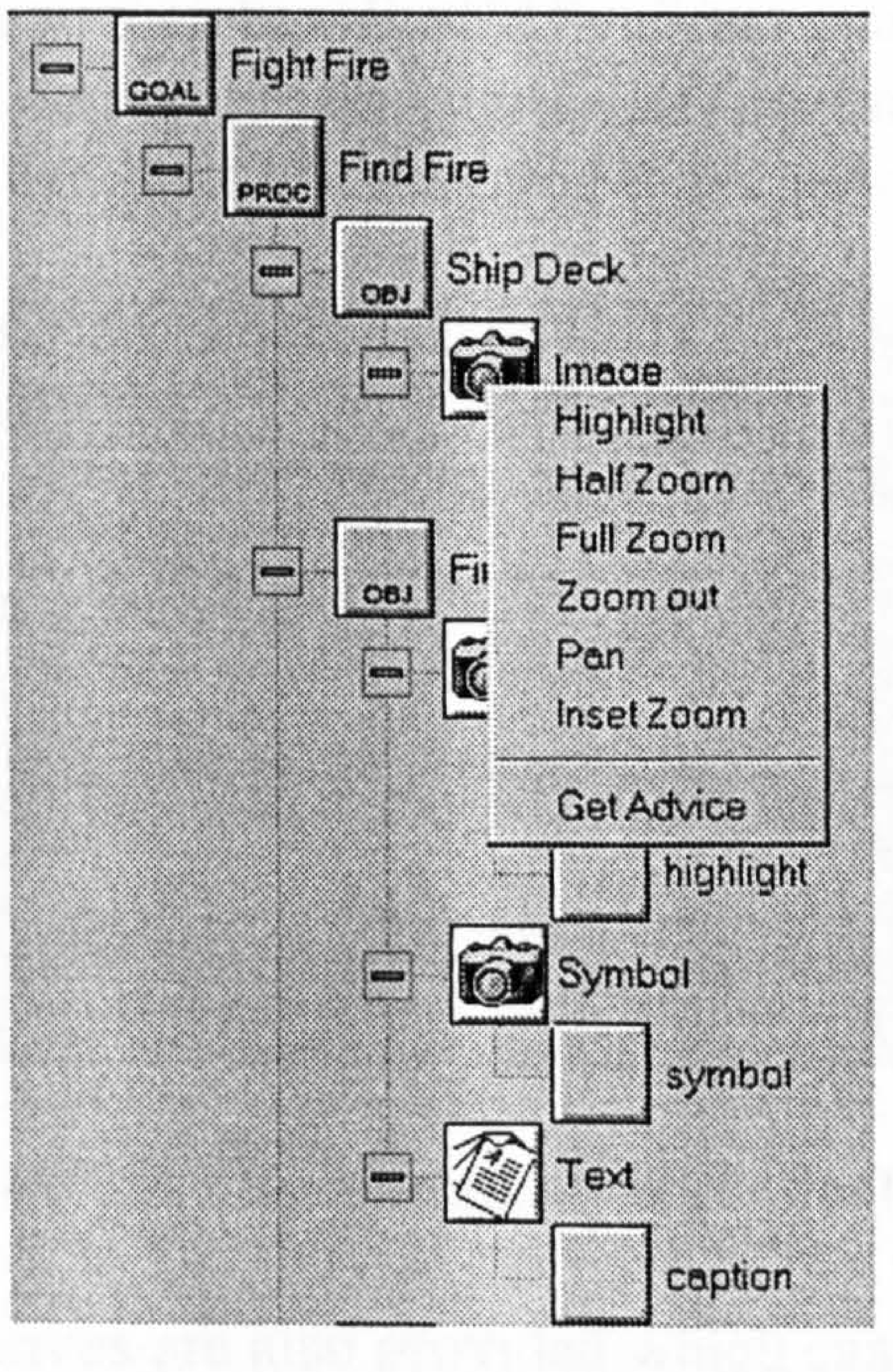


Figure 7.3.5 Adding an image presentation effect

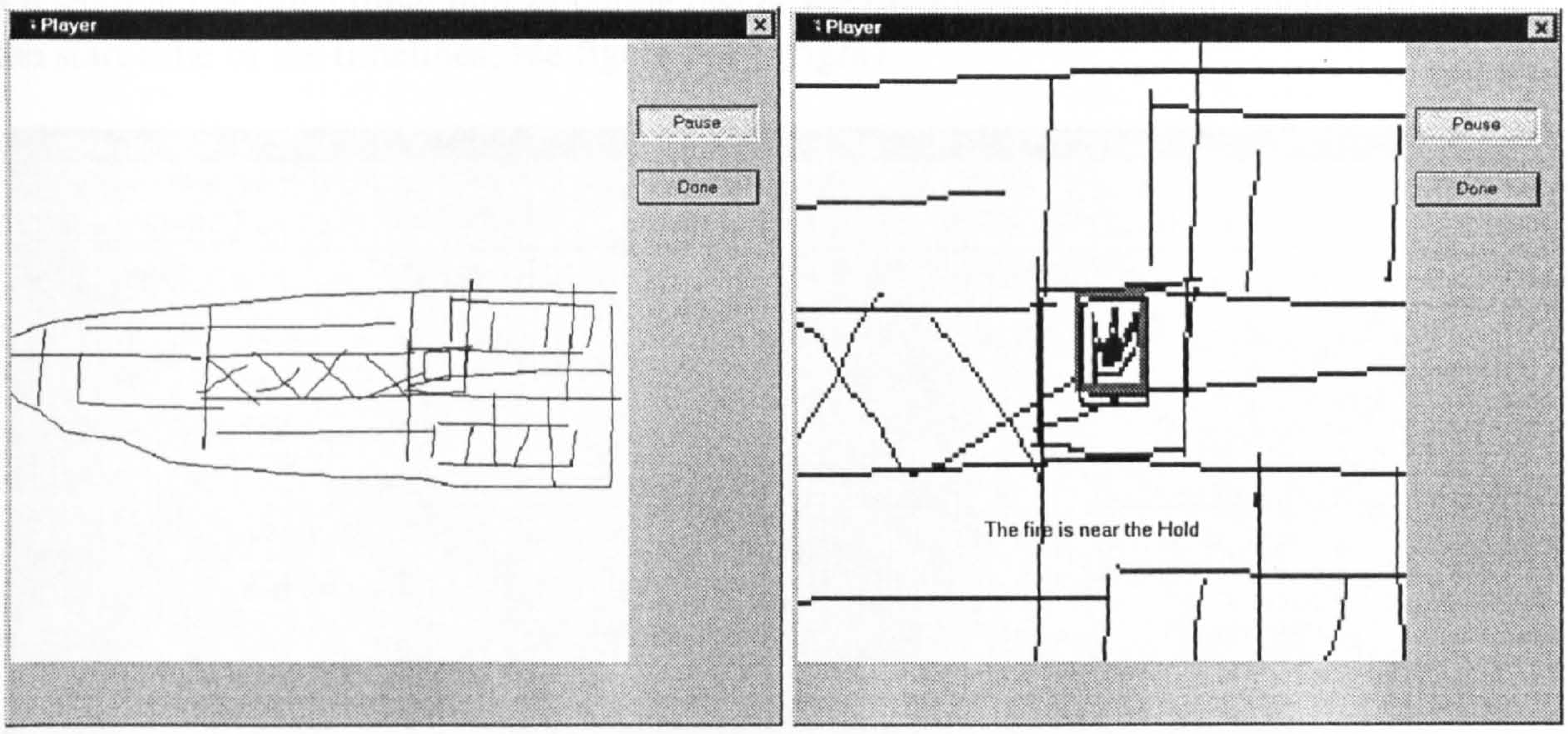


Figure 7.3.6 Example of still image (left) with zoom, icon, highlight and text caption effects (right).

7.3.6 Time Line Editor

The task-media model does not specify exactly how the presentation should be synchronised over time. To simplify the formation of the timeline, the task-media model is used to provide a first-cut sequence. All media resources and presentation effects are grouped in time so that they start at the start of the procedure they are associated with, and end at its end. Defaults are used for duration (see figure 7.3.7).

Since this initial sequence is unlikely to be useful to the designer (as everything would happen at once), the time line can then be adjusted by direct manipulation. The duration can be adjusted by dragging at the start or end of a particular timeline. The start and end time can be changed by clicking on the bar itself and dragging it to a new position.

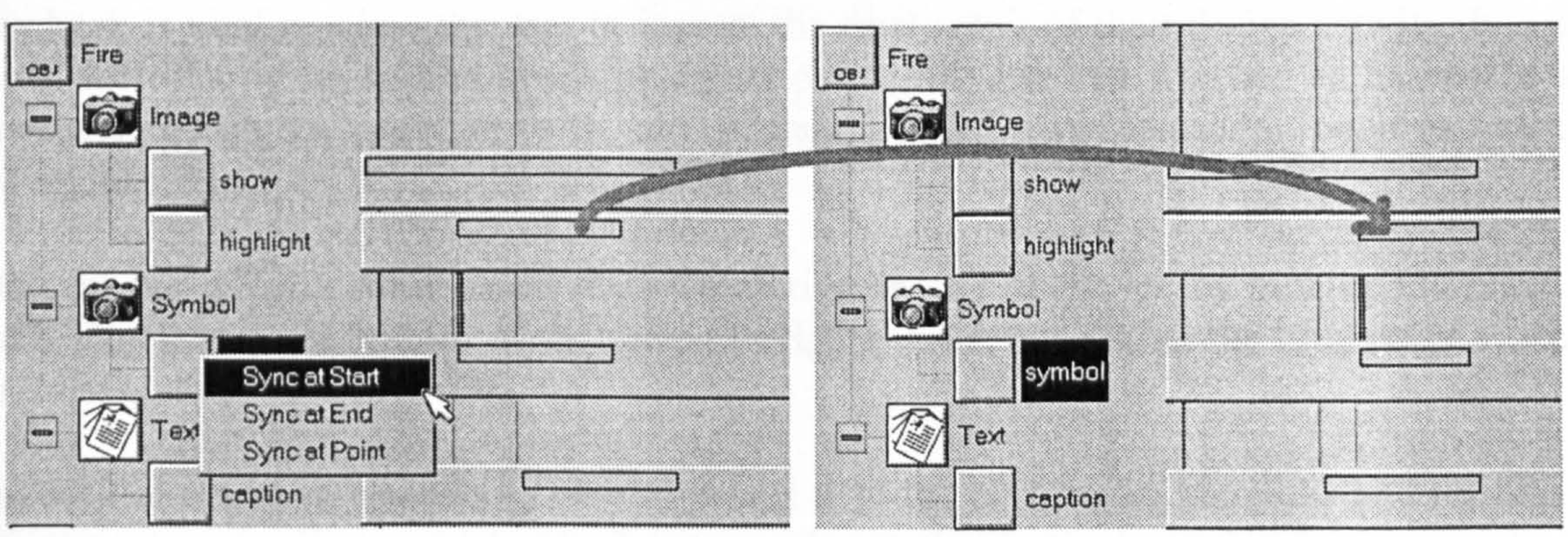


Figure 7.3.7 Synchronisation of timelines

A set of synchronisation primitives are also provided which can be used to structure the time line, so that a particular media resource or presentation effect starts or ends at the same time as another, or starts after a certain amount of time. These form constraints which are dynamically held upon the time line once specified. Thus two timelines can be constrained to e.g. start at the same time; thus dragging one of the timelines will then automatically update the start time of the timelines, see figure 7.3.8 (right).

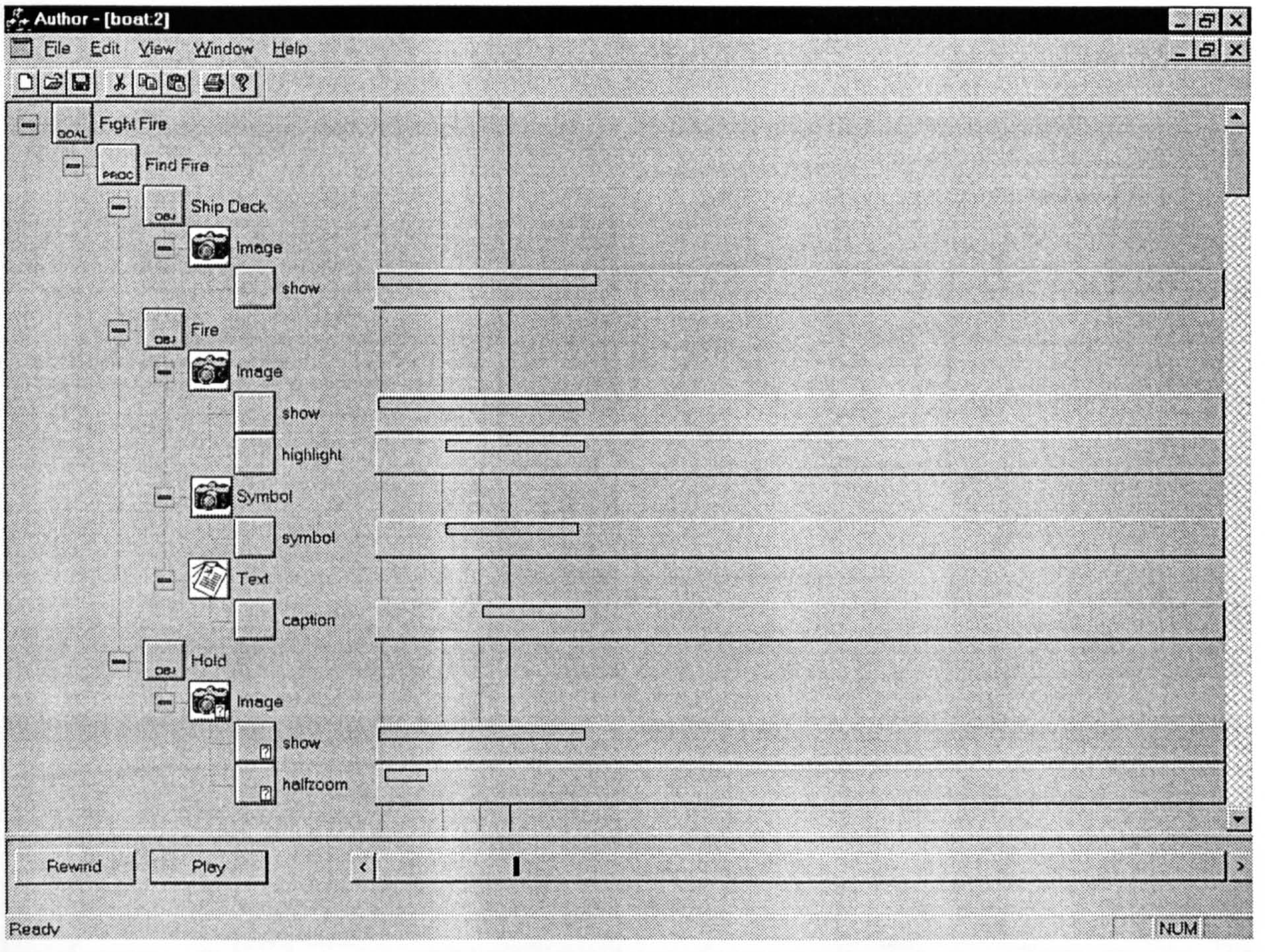


Figure 7.3.8 Timeline view

7.3.7 Presentation Play Back

Play back is vital to allow the user to inspect the final result of their design activity. It is also useful for directly specifying synchronisation points, the use can watched an animation and move the speech track around to fit with it, then set up a synchronisation point to maintain this relationship.

The user can specify from where the presentation should be played by moving the playback head. This lies at the bottom of the screen in figure 7.3.8, and shows the time point at which the presentation will play from.

Playback is achieved by making several passes over the task-media hierarchy and reading off the presentation effects which are due to render at the particular point in time set by the playback head. The playback head is then moved to its next position by the use of an internal interrupt driven timer. Figure 7.3.9 shows an example of playing a simple sequence containing a zoom, icon, highlight and label.

The presentation effects are rendered in two passes to ensure that scenes and objects are rendered before any viewpoint change is applied e.g. zoom, then any text captions are applied, so that they are not zoomed. Sound presentation effects are started and stopped when their timing points are reached.

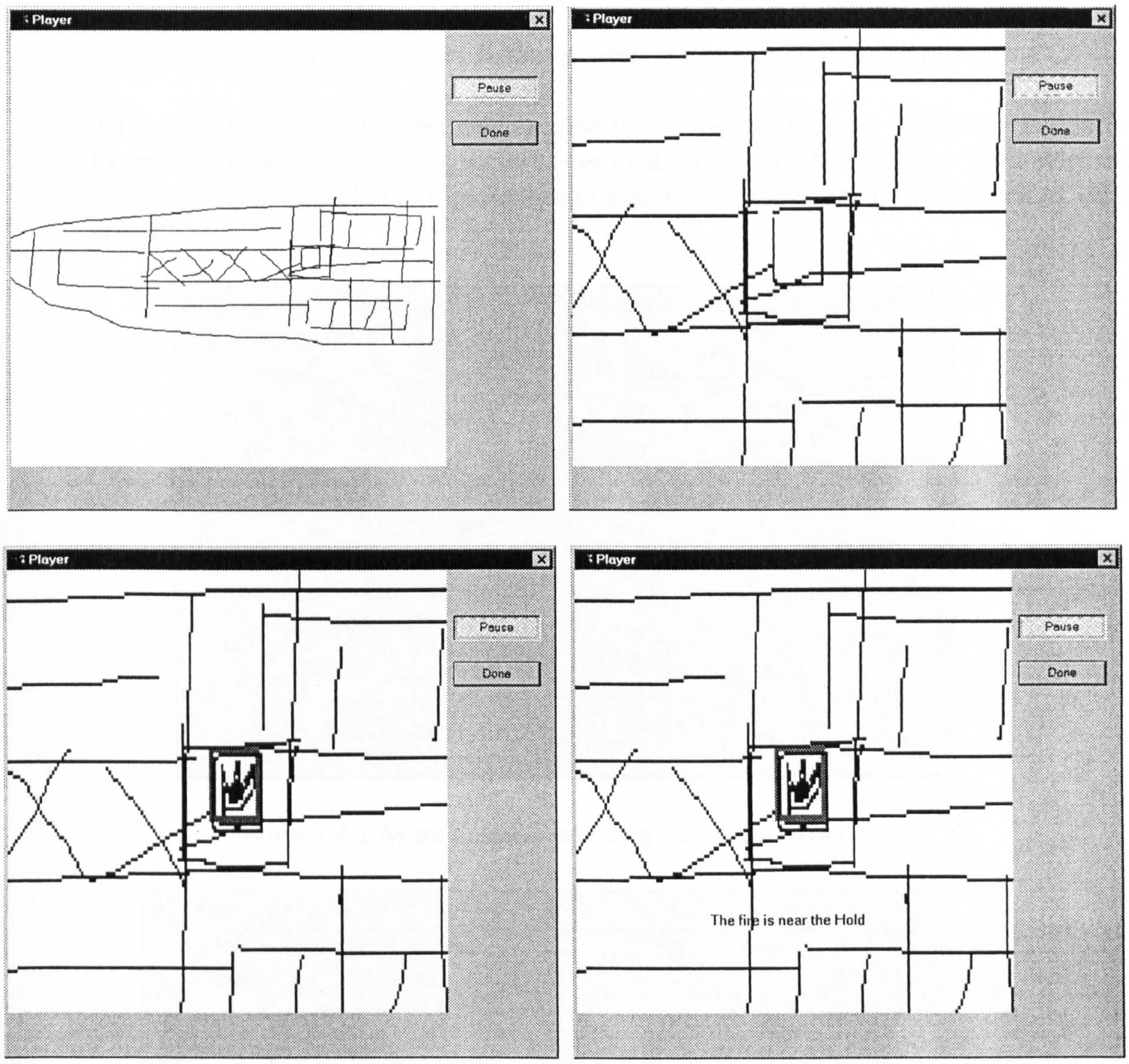


Figure 7.3.9 Example of playback for the find fire subgoal from chapter 2, showing an image with a zoom, icon and label effect.

7.4 Guideline support

7.4.1 Overview

The design advisor component uses the task / media representation and guideline rules to offer design advice and produce the attention graph. It was developed in prolog and provides a forward chaining inference engine which can load and interpret a set of rules from a text file. Prolog was chosen as it allows custom expert system shells to be easily prototyped. The main components are :

i) Guideline rules : the guidelines themselves are held in a free text file, written as a set of rules. The guidelines are divided into phases which represent the different steps in the method.

ii) Interpreter : this is a forward chaining inference engine, written in prolog.

iii) Author interface : this provides access to the media elements which are available for presentation, to the task structure, the media elements which have been selected for a particular task element, presentation effects and the start and end times of each effect.

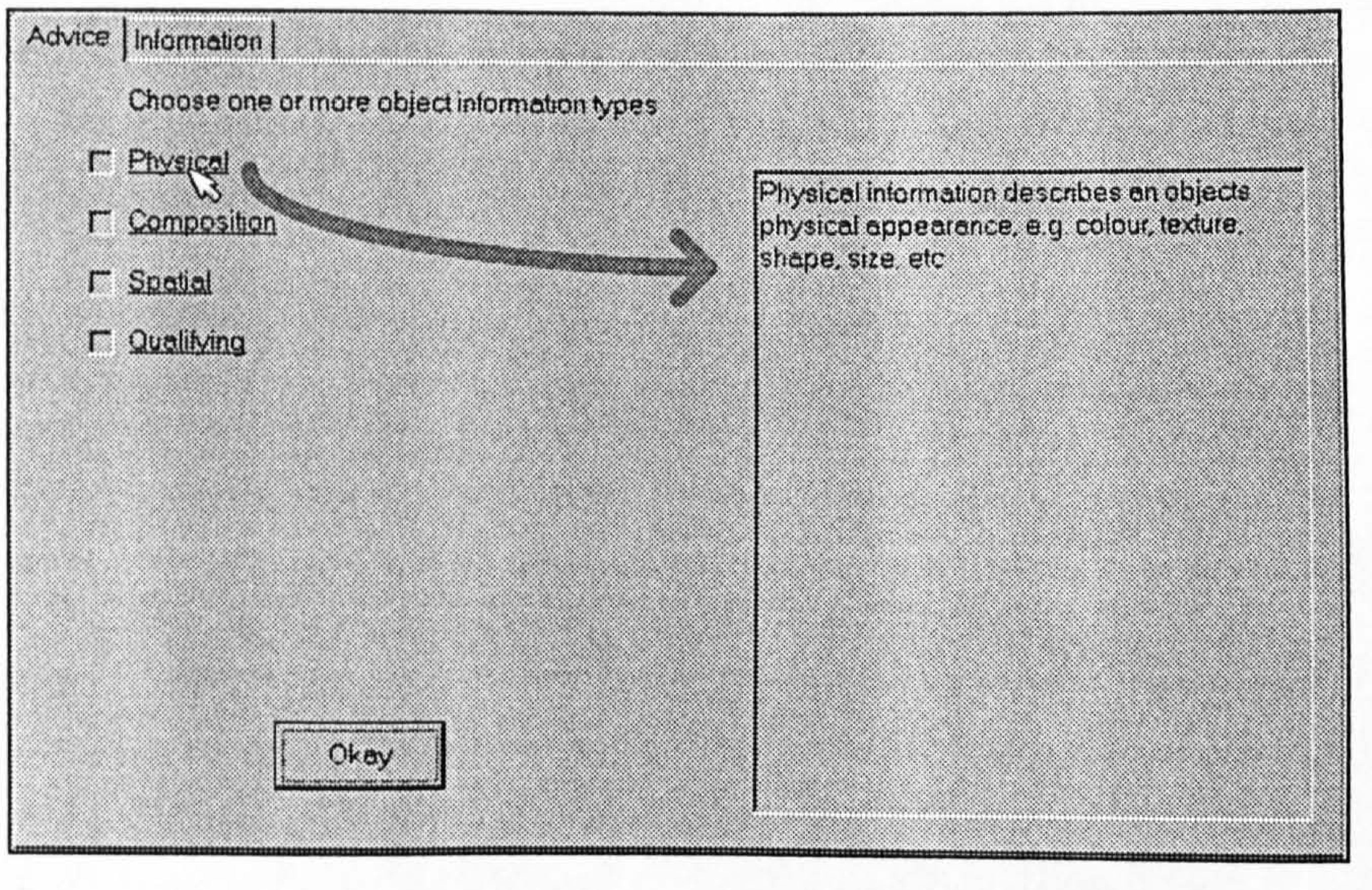


Figure 7.4.1 More detailed advice given from hyper-link

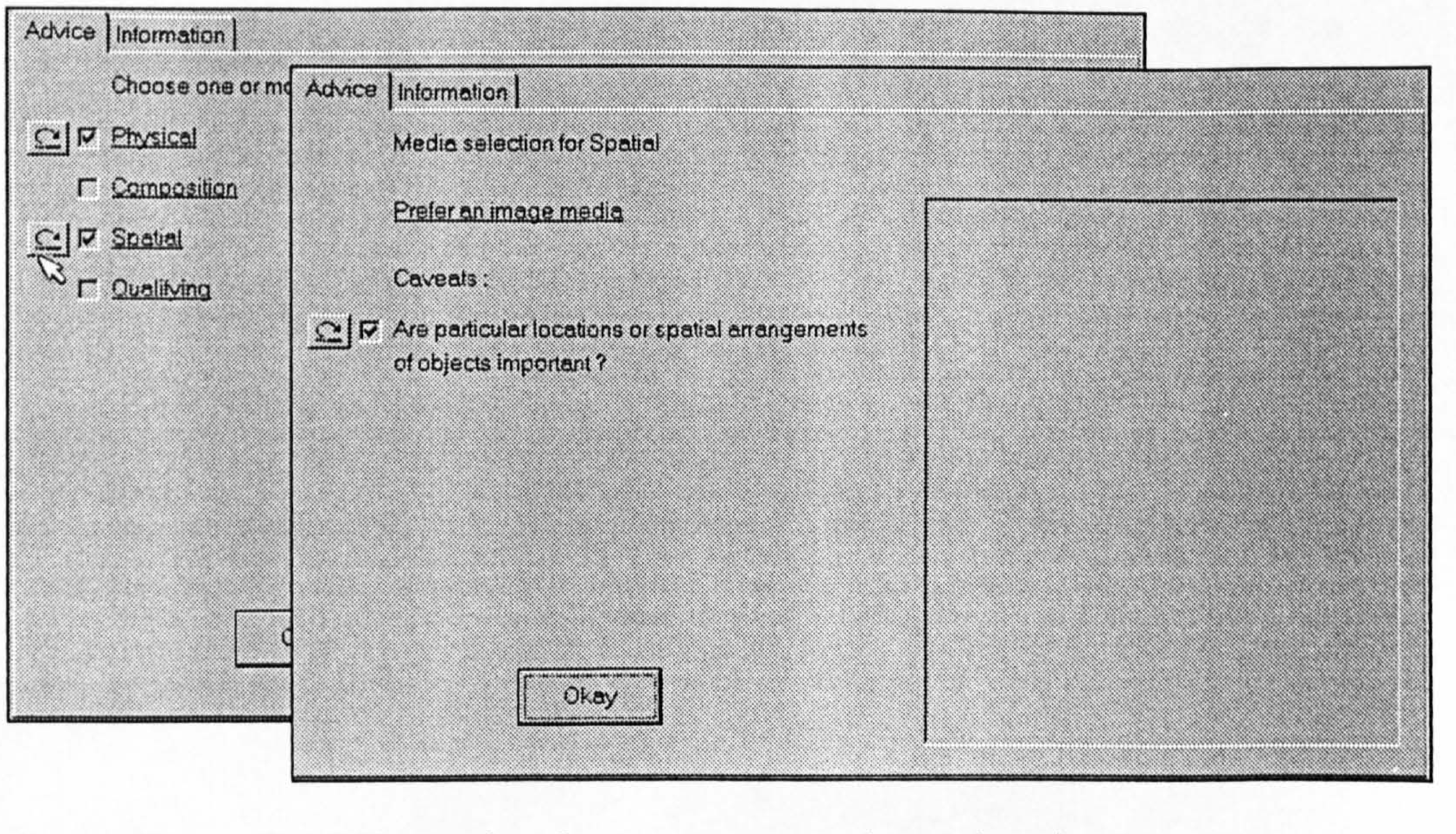


Figure 7.4.2 Example of expert system rules and review

The expert system is slightly different from a conventional expert shell. First, in order to support the task / media model, the working storage of the expert system is divided into a hierarchy of facts, each with a Node identifier. This hierarchy is updated whenever the user updated the task / media model. All facts are asserted with their node identifier. The system supports predicates to allow the hierarchy to be searched.

Second, it must support user interaction. This requires a set of predicates to provide menus, show advice, and produce the attention graph. The predicates also support a simple hypertext style display. Menus and advice text are entered as either a short summary form, or as longer text. The short text is given an underlined link, allowing the user to selectively view more detailed advice, or ignore it if they have read it previously. The current tool only allows a single short to long advice link. It is activated by moving the mouse over the underlined short text, causing the longer text to appear in the advice box on the right, see Figure 7.4.1.

Third, because the expert system is an advisor, it must allow the user to undo choices. This means that the system must allow inference to be restarted and the facts retracted when an undo is given from the user, and the facts re-asserted, minus any which would be invalidated by the undo (i.e. any facts which relied upon the undo fact). This facility is supported by separately asserting all the user's menu actions, and then re-running the inference following an undo, firing each rule via the stored menu options and only prompting the user for menu input if no previous menu input existed. Figure 7.4.2 shows a rule being refired to review it's advice.

Lastly, it is able to manipulate the task / media model. This is provided by predicates which access the model and add media or presentation effects. These are outlined in more detail below.

7.4.2 Task/Media Rules

The rules are made up of a left hand side which contains the condition which must be met in working storage for the rule to fire, and a set of actions which the rule triggers. Since the rules are separate from the authoring tool, they can be easily edited and customised to suit the guidelines.

The expert system uses two distinct knowledge bases of rules based on the guidelines. The first set deals with interacting with the user based on menus. These give advice on selection of media and presentation effects, and information required for the attention graph. To support the method, the rules are broken down into a two distinct groups, each with separate phases. The first group of phases deals with the task / media model: media selection, validation and presentation effect selection. The second group supports attention graph production and attentional validation.

The following sections detail the predicates supported by the expert system shell, and then provide examples of the rules used for each of the phases.

7.4.2.1 Rule overview

The task/media rules are defined in a number of phases :

a) Media selection : the media selection rules fire when the user asks for design advice on a task node in the task/media model. These rules use a menu offering the user a choice of information types, depending upon the task node the user has selected. The information types selected are asserted, firing other rules which narrow down the choice of media to be made. Finally, the media are added under the task node selected.

b) Validation : after each media selection rule is fired and a new media added to the task/media model, a set of validation rules are fired. These check the current set of media under the task node for any combinations which may require particular design care.

c) Presentation Effect selection : the presentation effect rules fire when the user asks for design advice on a media node. These rules offer the user a choice of presentation effects, depending upon the media node the user has selected.

7.4.2.2 User interaction predicates

The prolog interpreter supports a range of extended predicates to allow it to interact with the user :

- menu(fact(X), Text, [fact(X), Text]) : where the user can choose from a set of options shown as a menu. Each question instruction causes the text accompanying the question to be shown as set of check boxes. The users selections are asserted in working storage, causing more rules to fire

- write(Text) : allows text to be displayed as advice to the user.

- wait : causes the expert system to wait until the user presses the okay button.

- showlight(Node1,Node2,State) : shows a warning light on a particular node in the task / media model

7.4.3 Media and Task predicates

As noted earlier, media and task-media elements are both stored in hierarchies of nodes, allowing the media and task to be defined. The following instructions are supported :

- has(Node,fact(x)) : the expert system can recursively search over the task-media nodes; this allows the child nodes to be searched for particular facts.

- this(X) : provides the current task-media node. The node can interrogated to find out which task-media node the user has currently selected in the task-media hierarchy. The 'set type' and 'check type' instructions both operate upon the node held in this variable. The node pointed to can be updated using the 'for each node' instruction.

- addmedia(Node,Media)) : the expert system can add a new media node to the task node given in Node.

- addtech(Node,Tech) : the expert system can add a new presentation effect node to the task node given in Node.

7.4.4 Example rules

7.4.4.1 Media selection example

The following shows an example of a rule to allows the user to provide details of a particular information type. The results of the rule firing are shown in figure 7.4.3.

```
rule selection-comp2:
  [  infotype(composition)]
 ==>
  [  write('Media selection for Composition'),
     write(''),
     write('Use a visual media','If the activity requires
composition information then use   a visual media'),
     write(''),
     write('Caveats :'),
     menu([    composition(abstract),'Is asbtract information
being explained ?',
          'Abstract information is no-physical object details,
such as number or type',
          composition(granularity),'Is the granularity of the
components important ?',
          'Granulairty information concerns how the object is
composed as a whole from          one or more parts'])
  ].
```

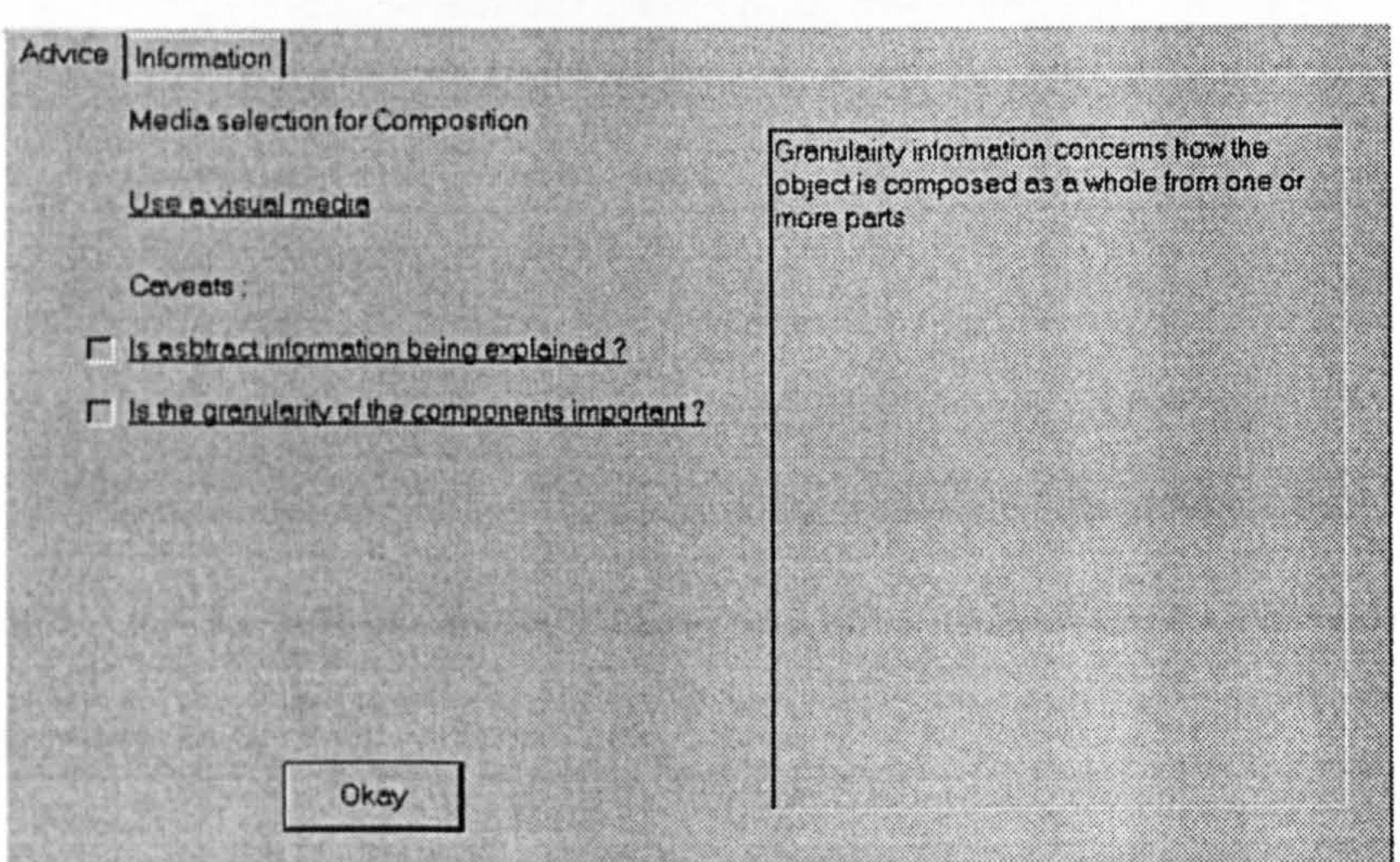


Figure 7.4.3 Example of menu rule and screen shot of menu produced.

7.4.4.2 Effect selection example

The example rule below displays a menu which prompts the user for information concerning the visual appearance of an image. When the user selects one of the menu options, either image(emphasised) or addtech(image) are asserted causing further rules to fire to help the user choose a presentation effect. This is shown in figure 7.4.4.

```
     rule selimage:
        [  selection,
           type(image)
        ]
       ==>
        [  write('Guidelines for Image Design'),
           write(''),
           write('If the object shown in the image is important
     then either :'),
           write(''),
           menu([     image(emphasised),
                      'The  image  is  large,  detailed,  bright,  away
     from rest',
                      'These physical features make an object likely
     to be focused on by the         viewer',
                      addtech(image),
                      'Add a presentation effect',
                'Presentation effects include highlight, label and
zoom. These will                    help attract attention to the
image.'])
          ].
```

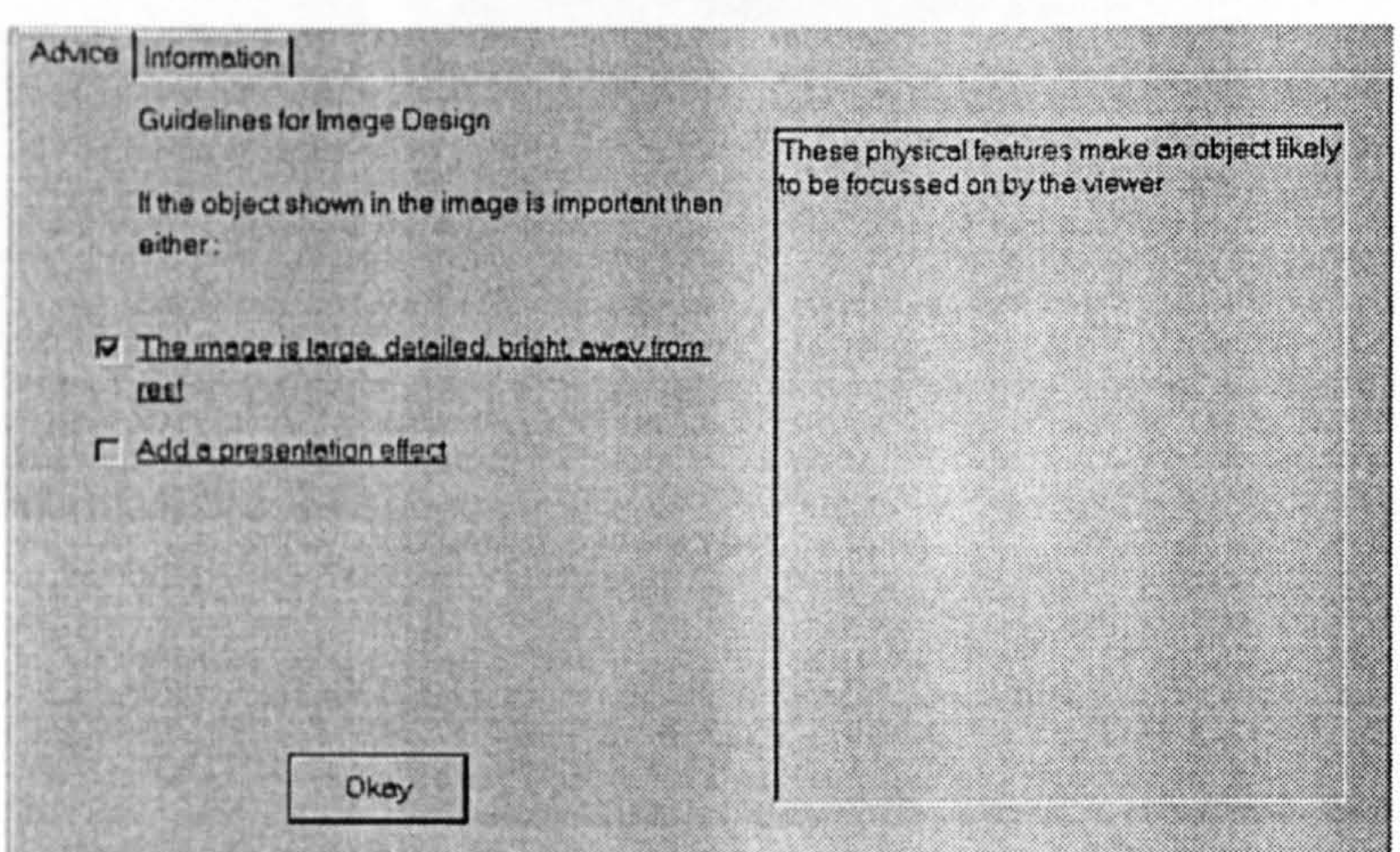


Figure 7.4.4 Example of menu rule and screen shot of menu produced.

7.4.4.3 Validation example

A validation rule is shown below. It is used to check that speech and text contain the same content. First it uses not(valid(labelsp)) to make sure that the warning for this task node has not already be given, then it interrogates the media nodes under the task node using this(X) to find the current task node and has(X,Y) to find the facts that the media contain. The and(Clause,Clause1) predicate is used to perform a backtracking search using prolog. Finally, showlight(Node,Node1,State) causes a light bulb icons to be lit up in the task / media model showing which nodes are being referred to. This is shown in figure 7.4.5.

```
rule selection-validlabelsp:
[     dovalid,
      not(valid(labelsp)),
      this(X),
      has(X,Y) and node(Y,type(speech)),
      has(X,Z) and node(Z,type(text))
  ]
 ==>
  [   assert(valid(labelsp)),
      showlight(Y,Z,1),
      write('Warning'),
      write(''),
      write('Validation rules for text and speech'),
      write(''),
      write('Make sure wording is similiar for text and
speech',
      'When using a label and speech the wording and content of
the text and speech should   be as similiar as possible'),
      wait,
      showlight(Y,Z,0)
  ].
```

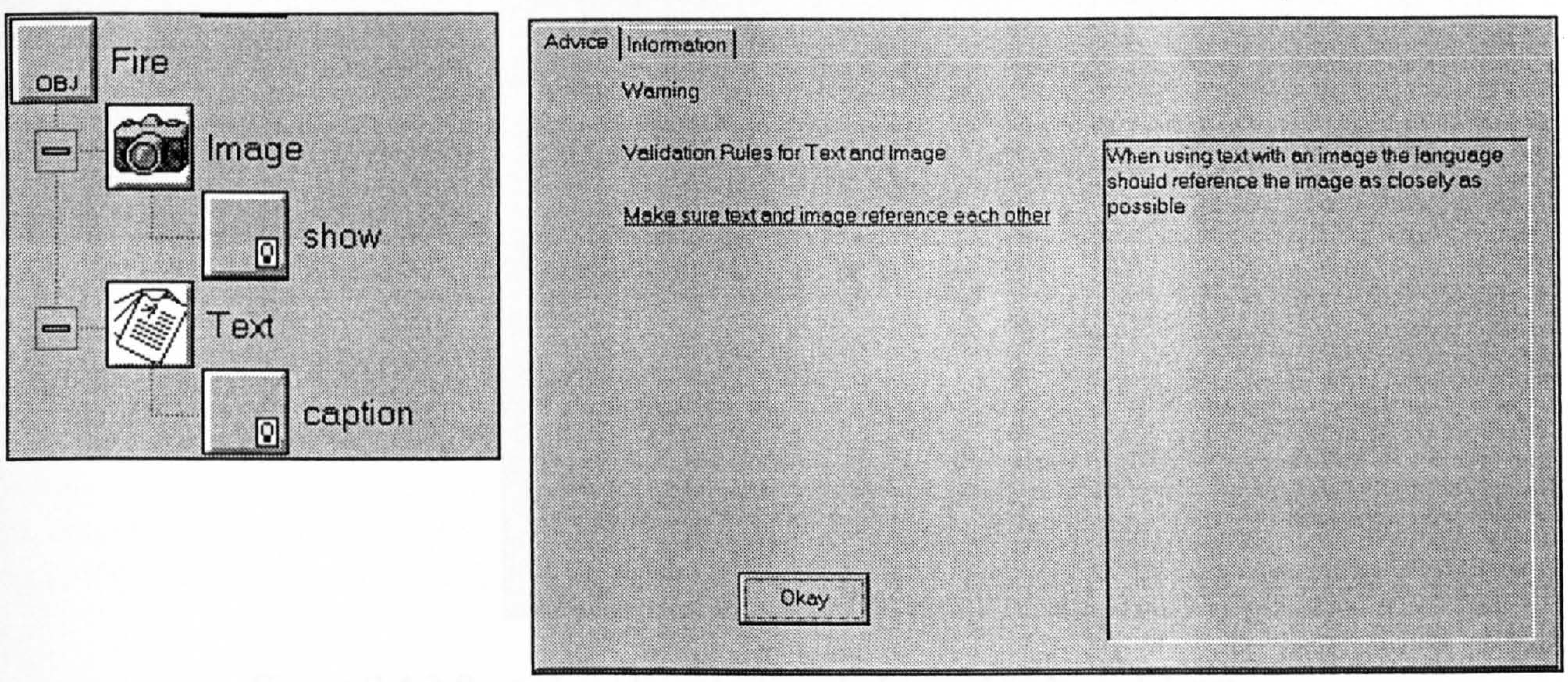


Figure 7.4.5 Example of validation rule and screen shot of menu produced and light bulb icons in the task / media model.

7.4.5 Case Study Example

The case study example is based upon the 'find fire' subgoal from the shipboard emergency task. The following screen shots illustrate how the expert system advises on designing the media for the 'fire' object. In figure 7.4.6, the user selects the relevant information types. By moving the mouse over the information types, more information can be viewed.

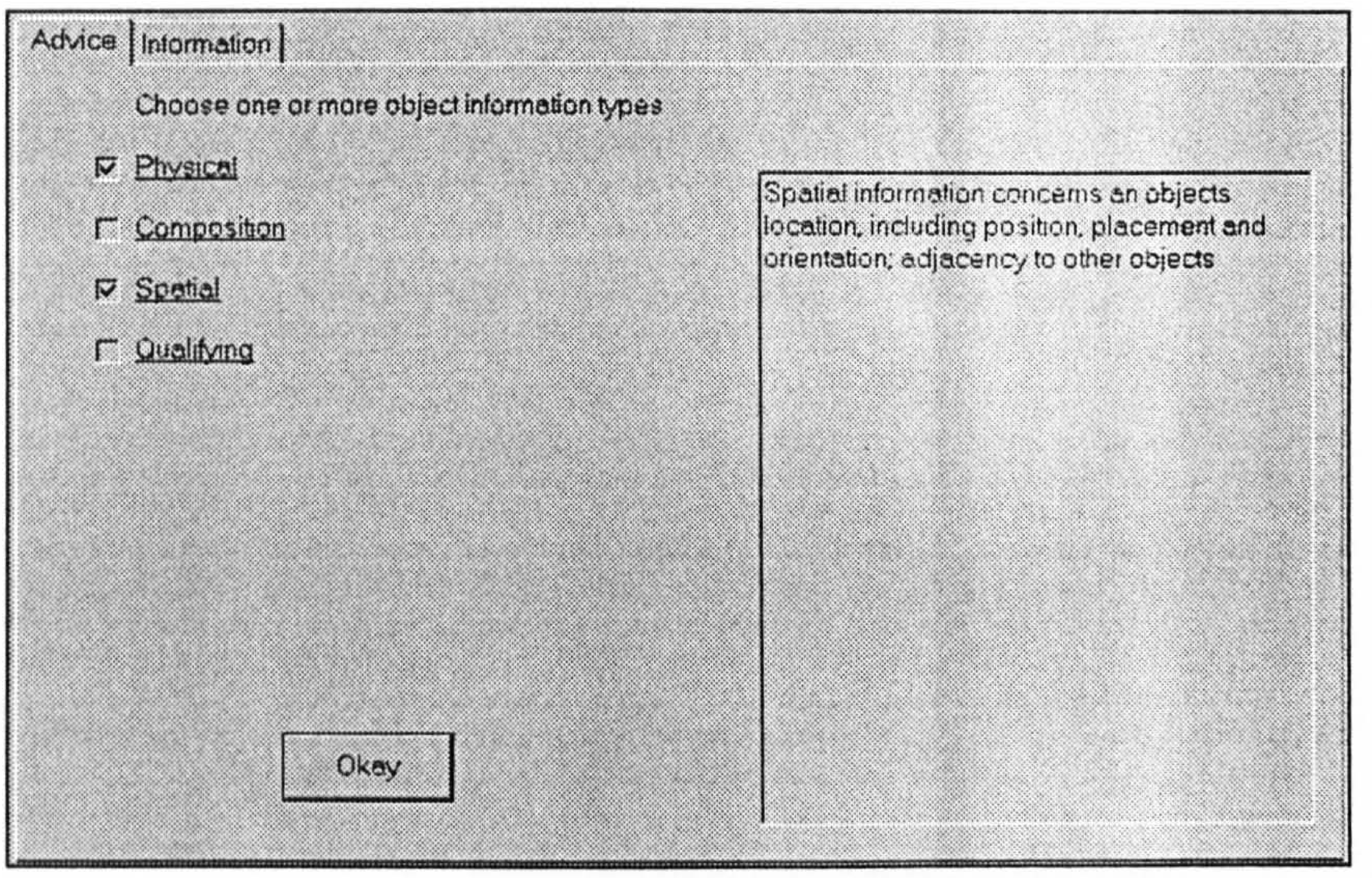


Figure 7.4.6 Media selection dialog

Having selected physical and spatial as being important information types, the expert system now provides advice for media selection. It first suggests an image media should be used for spatial information. If no image media is currently present, then it adds an image to the task node. This is shown in figure 7.4.7.

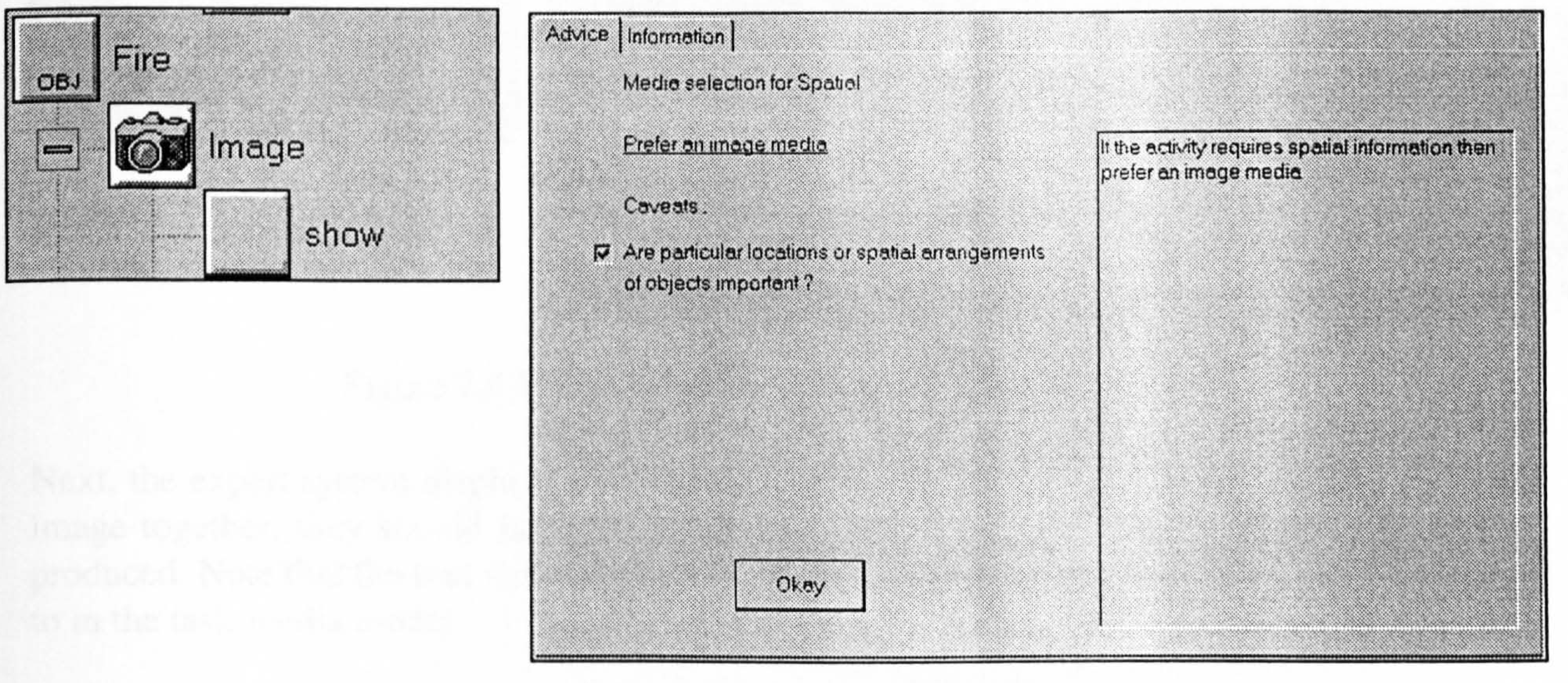
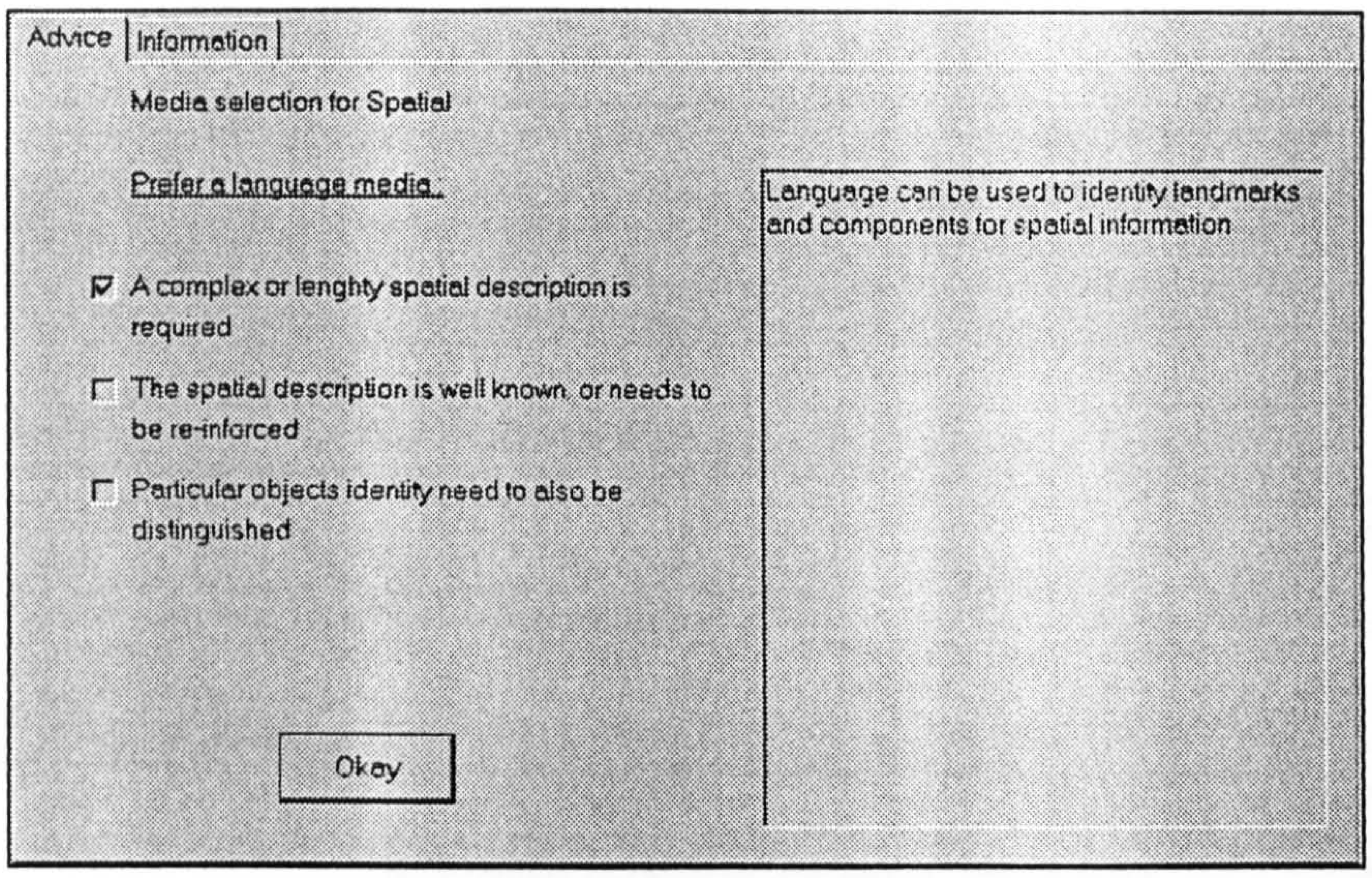


Figure 7.4.7 Spatial media selection dialog and task model state

The user then selects that additional spatial information is required. The expert system asks what type of detail is required, and the user selects that the spatial information is complex. The expert system suggests using a text caption, and adds one to the task node, see figure 7.4.8.



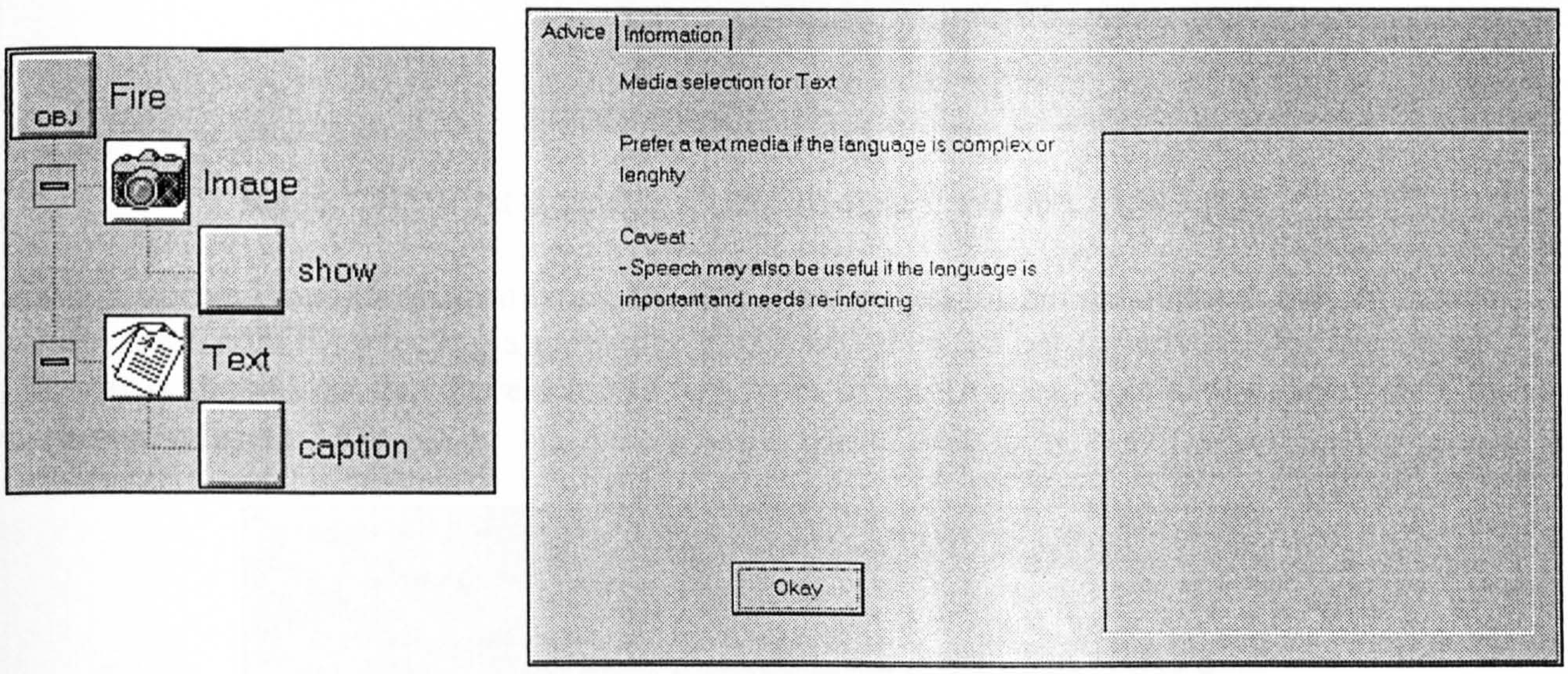


Figure 7.4.8 Text selection dialog and task model state

Next, the expert system displays a validation warning. It suggests that when using a text and image together, they should have the same content. Figure 7.4.9 shows the warning dialog produced. Note that the tool shows light bulb icons to highlight the nodes which it is referring to in the task-media model.

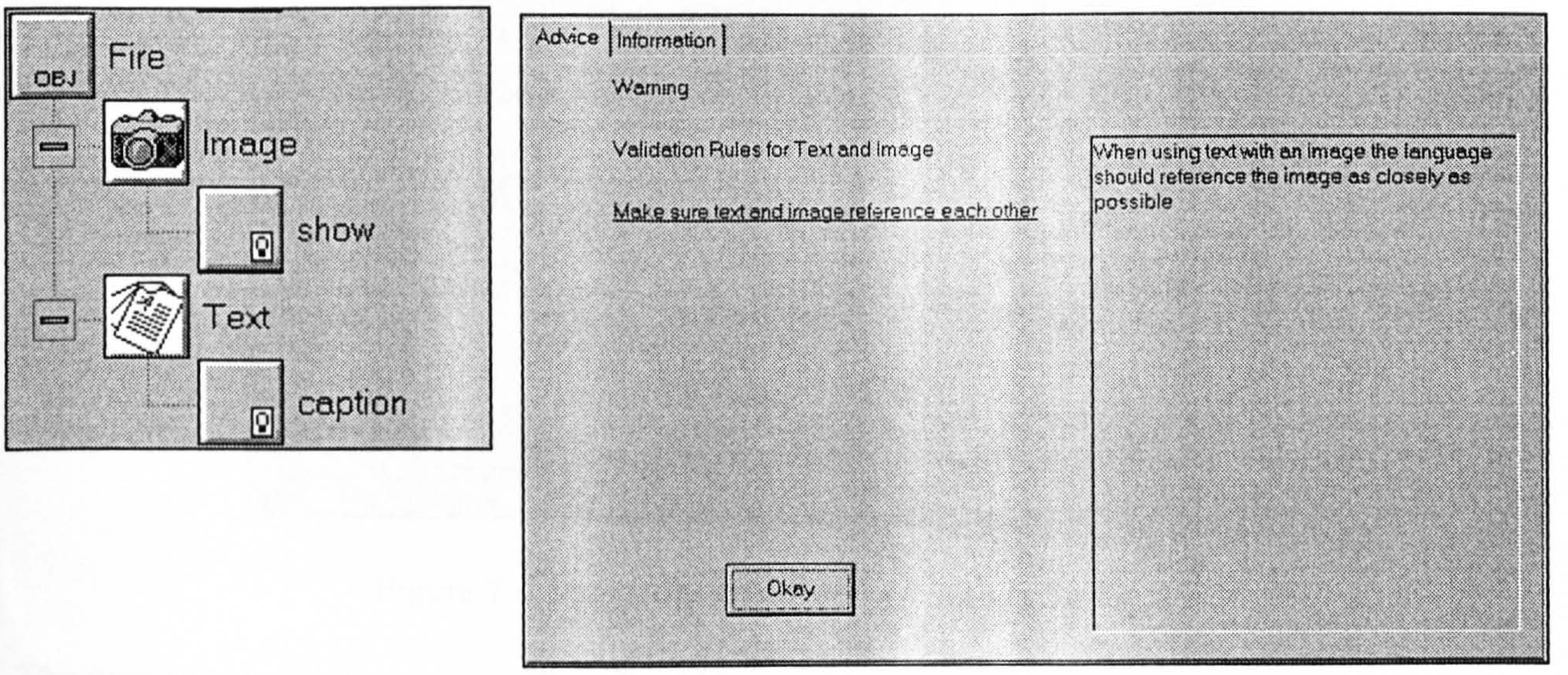


Figure 7.4.9 Text & Image validation dialog and task model state

Having selected the information types, the user can then get further detail about how to design within the media by selecting a particular media and then asking for advice. For the image, the user is shown that they have selected two types of information, spatial and physical. This is shown in figure 7.4.10.

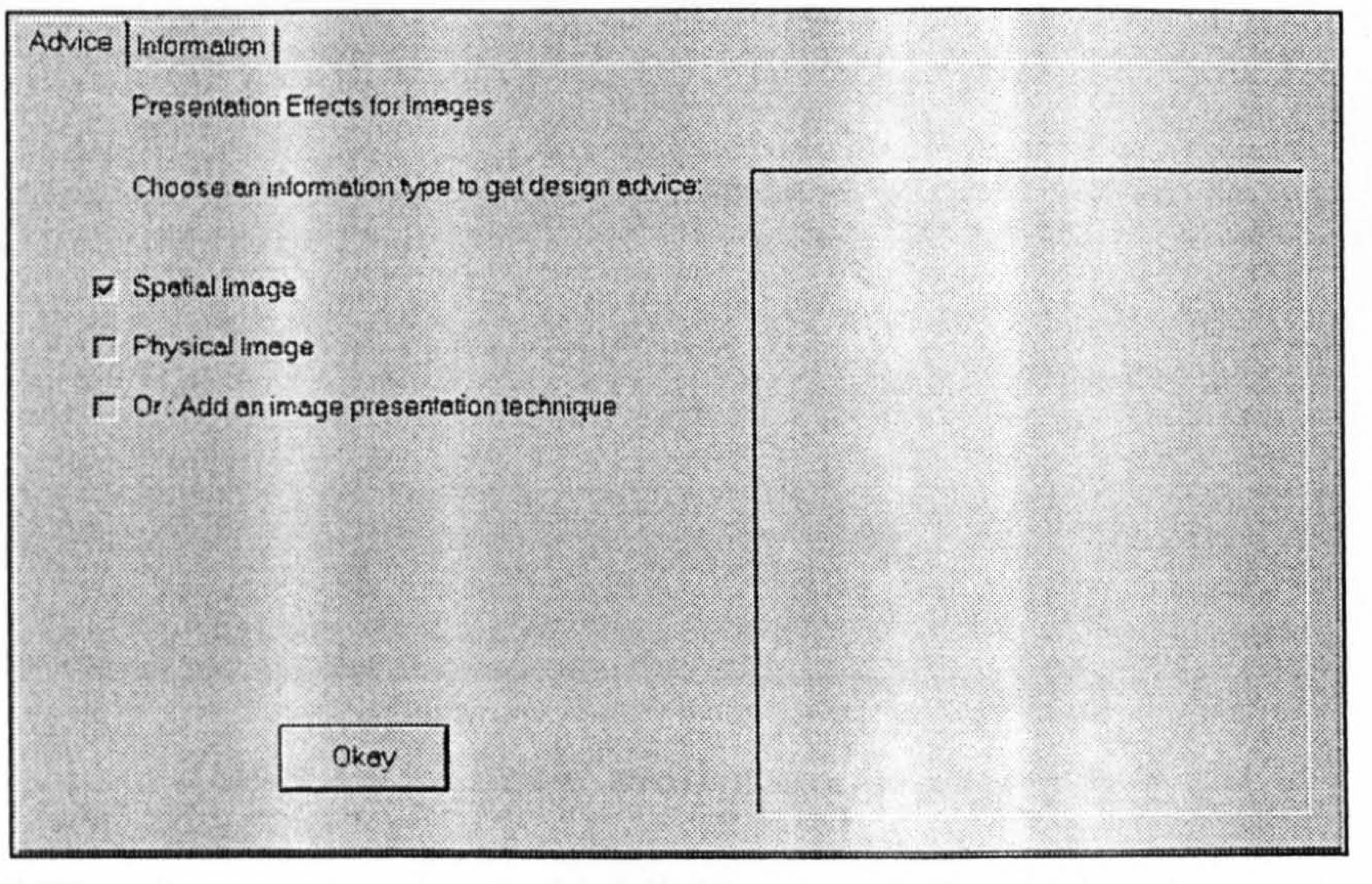


Figure 7.4.10 Within media design dialog

The user opts to get further advice on designing spatial images, shown in figure 7.4.10. The expert system then gives information concerning how to design landmarks, and suggests adding a presentation effect if the object is important, see figure 7.4.11

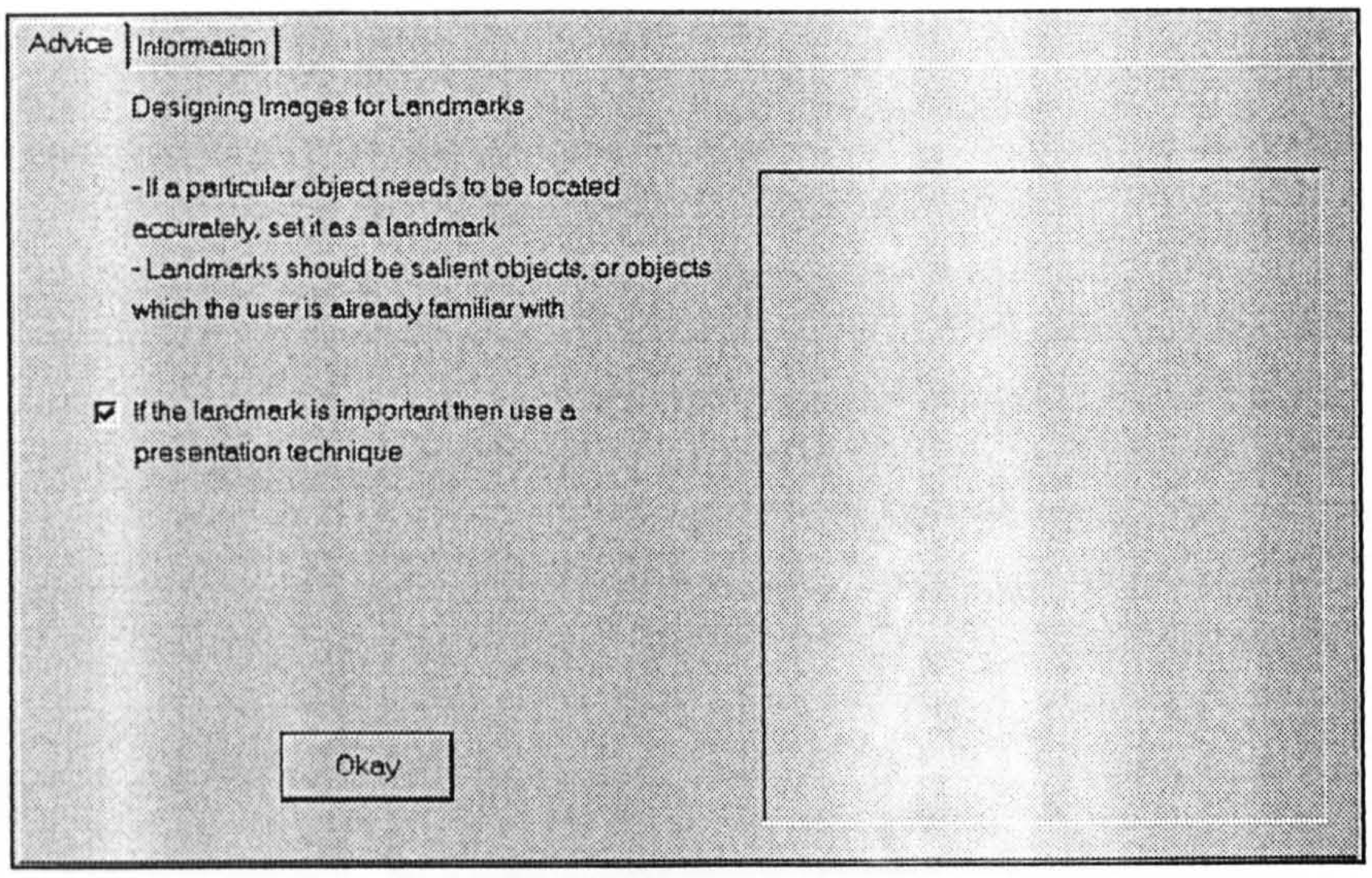


Figure 7.4.11 Spatial image design advice dialog

The user selects a presentation effect should be used because location of the fire is important. The expert system then asks further questions about the nature of the image to decide which type of effect should be used. The user selects that the object is small and may be missed, see figure 7.4.12.

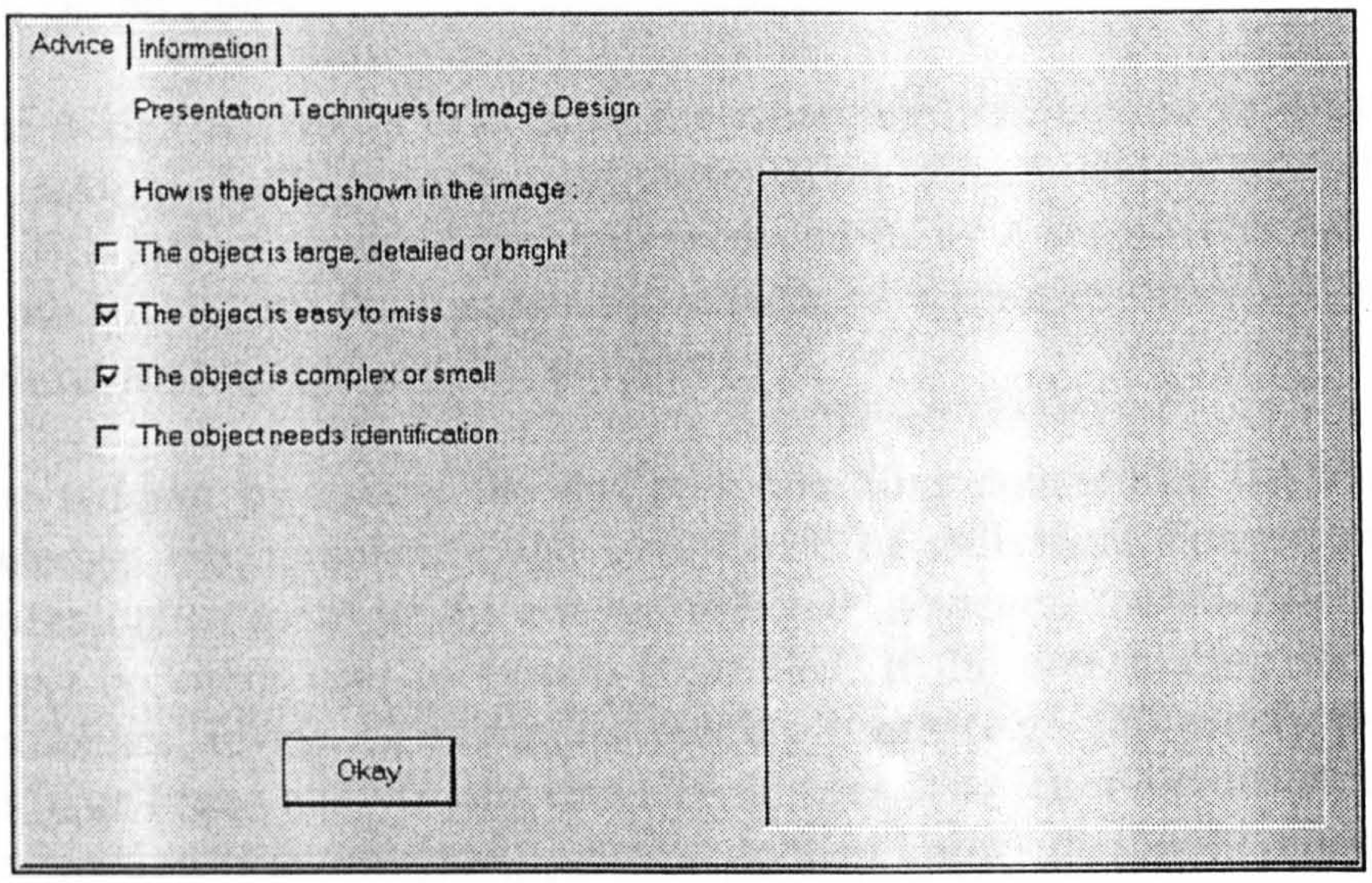


Figure 7.4.12 Image presentation effects dialog

The expert system advises on using a highlight and symbol effects to aid the viewer in locating the fire. If these effects are not already present on the image media, then the expert system will add them. Dialogs are shown advising on the design of the symbol.

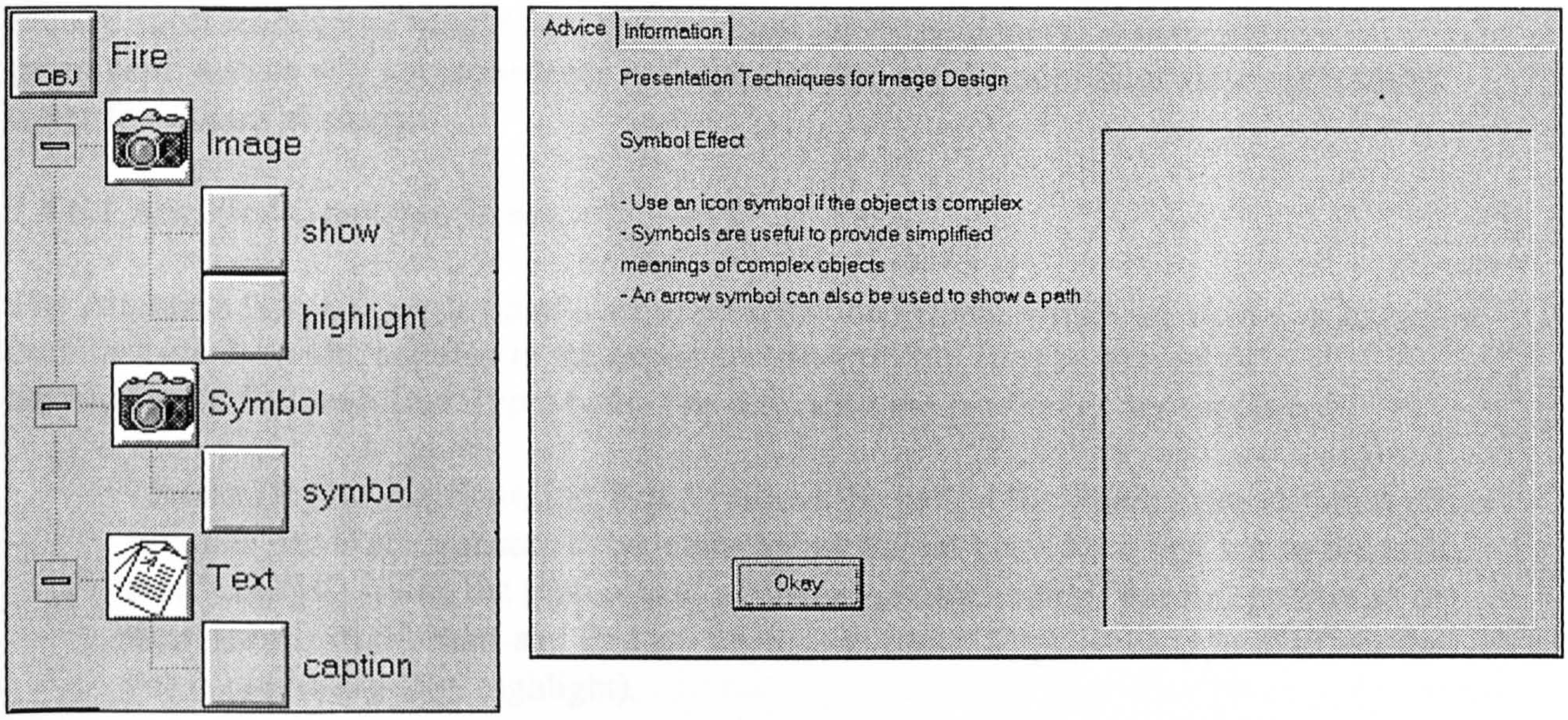


Figure 7.4.13 Symbol presentation effect advice and task model state

The final state of the find fire node is shown in figure 7.4.13. It now has text, image and symbol media, with a highlight effect on the image.

7.4.6 Attention Graph Rules

The aim of the design advisor is to critique the emerging presentation as the user builds up the model by adding media components, presentation effects and adjusts timing via the timeline bars. The 1st cut attention graph is produced using a set of rules which search the task / media model looking for presentation effects in a priority order and adding default assumptions about reading and viewing durations.

These may be tailored by asking the designer questions concerning the user's profile of domain knowledge, the complexity and importance of individual images and labels. The duration of speech clips is set by the content selected; however, the display duration of text and images may be either user or system controlled. If the presentation sequence is to be controlled guidelines advise on minimum reading times (e.g. 120 words per minutes for simple text in large fonts) and viewing times for images. The latter are difficult to estimate as information extraction is determined by users' domain knowledge, motivation as well as by image complexity.

The rules search through the task / media model finding presentation strands that are concurrent, then identify synchronisation problems and potential conflicts in attention that may arise leading to the display of warnings with design advice. The notion of a 'critic' is similar to [Fischer et al 90] who defined critic based advisors that do not automate design, but attempt to spot and alert users to design problems which they can then fix.

The tool allows limited adaptation of the critics and the attention graph. The designer can choose to de-activate any warnings with which they disagree. The tool also allows the designer to selectively add information concerning the media to tailor generation of the attention graph. Currently, the designer can specify that media are familiar, and thus may require short reading/viewing times; that an image (or a component) is perceptually prominent, and so will get precedence in focus; and can set the minimum viewing/reading times for a class of media.

7.4.6.1 Attention Graph predicates

The Attention Graph is made up of a set of presentation effects, which are held as a list upon each task-media node, asserted in the expert system working storage as a set of hasbar(Parent,Bar,Start,End,Type) facts. The following are predicates are supported :

- hasbar(Parent,Bar,Start,End,Type) : allows the period for which a presentation effect timeline bar is set in msecs to be interrogated. Parent provides a link bar to the task / media model, giving the type of task node for the timelinebar. Bar is a pointer to the timeline bar itself. Start and End are the time in msec, Type is the type of effect the bar is producing (e.g. highlight).

- overlap(Start1,End1,Start2,End2) : allows the list of all presentation effect which have a period of overlap to the current presentation effect node be iterated over. This uses prologs backtrack search facilities.

- setfocus(Bar,Start,End) : set a focus bar for the current timelinebar given in Bar, with Start and End as times.

- hasfocus(Bar,Start,End) : returns the current focus Start and End times for a timeline bar

- addwarn(Bar,Text,Time) : set a warning button on a bar for the current timelinebar given in Bar at the time position given in Time. When it is selected by the user, it will display the text given in warn.

7.4.6.2 Rule overview

The rules used to produce the attention graph will now be described in more detail. They are divided into 6 phases :

a) Initialise : The first set of rules are used to initialise the focus search. Housekeeping rules check that focus can only be set on highlights, labels and animation when their associated media component has been displayed. The rules also create separate timelines for visual and audio media.

b) Find a candidate start time : The next set of rules search through the start and end times in the media model, seeking components which could potentially be in focus. The start time of one of the components is selected, based on the following salience priorities.

> i) Any animation will be selected first. These rules recognises that animation have a very high attentional value and will gain focus as soon as it is released by other salient

effects, and will act as an interrupt on media with lower salience, e.g. text and still image.

ii) Attention is set to highlights when they are displayed. Highlights have a lower priority than animation with an inbuilt warning factor of 200 ms, so other effects can compete for attention soon after the highlight is displayed. The tool makes provision for multiple highlight types, which can be set to different priority levels (e.g. colour icon, box surround, point by arrow).

iii) Attention is set to labels when they are displayed. Revealing labels has a lower salience than animation or highlights, and the effect wanes rapidly so other effects can complete for attention soon after (200 ms default) a label is displayed.

iv) The start time is set to media components in the absence of any salient effects or animation. Image is selected as a default over text, while other user configurable rules can discriminate between media components on criteria of window size, position, image complexity, and text formatting characteristics (size, fonts, etc.). The advisor can only work with this information if it is input by the designer, and at present does not use this level of detail.

c) Check visual focus priority : Having found a candidate timeline start time, the advisor now prioritises the presentation effects throughout the script using the following rules.

i) If more than one effect can start at the same time, effects are favoured which are semantically related to the prior focus; this ensures that the current focus will move e.g. from an image to the image's highlight or label effect.

ii) Focus is given to highlight or labels which are also referenced by speech.

iii) As salient effects finish or time-out, attention is set to any effects or components which are available, i.e. viewable by the user, but have not yet gained focus. The rules backtrack over the timelines searching for effects which did not gain focus earlier, using priority rules and then recency to adjudicate between effects with equal priority.

d) Set focus duration. These rules set the focus duration to allow a minimum reading time for a label/text or viewing time for an image or animation.

i) Animations are given focus for their entire duration. This is because tracking motion requires focus to be maintained.

ii) Labels are given 200 msec focus per word in the label; giving a reading rate of 2 secs for a simple sentence

iii) Longer texts are given defaults calculated at a 120 word per minute reading rate.

iv) Single, simple components in images are given 200 msec for focus.

e) Add warnings : A set of critic rules are applied to spot potential focus problems. These step through the task / media model testing which media components and effects may compete for the user's attention.

i) A search is made for any effects with overlapping durations which would compete for focus. These rules find any effects which have been prevented from getting focus because of competition from a dominant effect. The rules add warnings and design advice e.g. if an image is revealed during an animation, a warning will be given because the animation will block focus on the image.

ii) Inter component and effect dependencies are checked, e.g. if a label is shown, then the referenced image component is checked to ensure that it is viewable and that attention to it has not been blocked.

iii) A search is made for speech effects which are within the same task procedure as text labels. The system adds advice that labels should be shown when cued in the speech.

iv) Finally, the media model is checked to ensure all parts of the message have been presented, and that attention to each component has not been over-ridden by other components with more dominant effects.

7.4.6.3 Case study Example

The screen shot below shows an example of how the rules produce an attention graph for the find fire subgoal. The white lines show the rules prediction of which part of the presentation will be in focus, and for how long. In the examples shown, simple image components require at 200msec viewing time, whilst text is given 5 words per second viewing time.

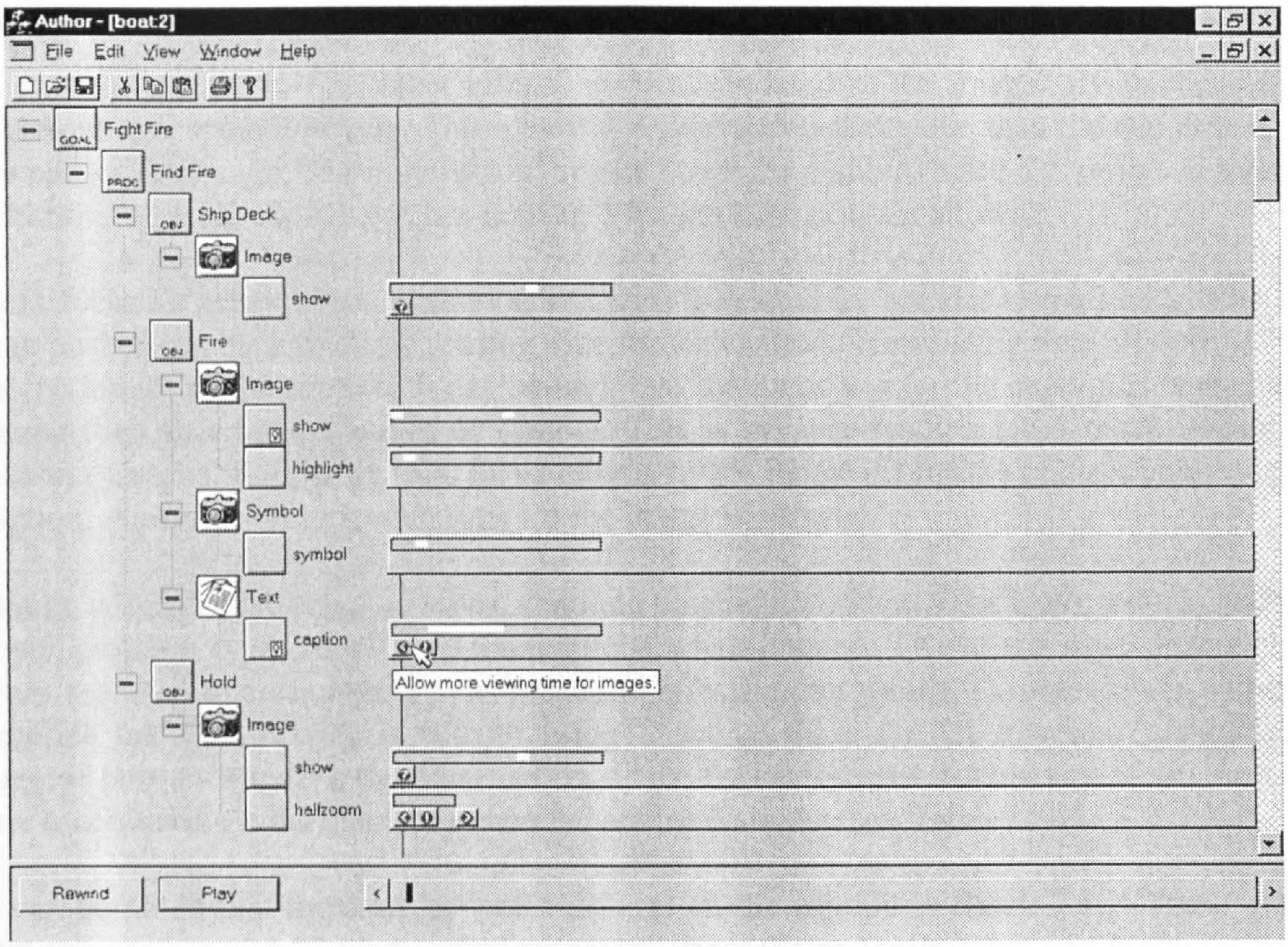


Figure 7.5.1 Default arrangement for find fire sub-goal

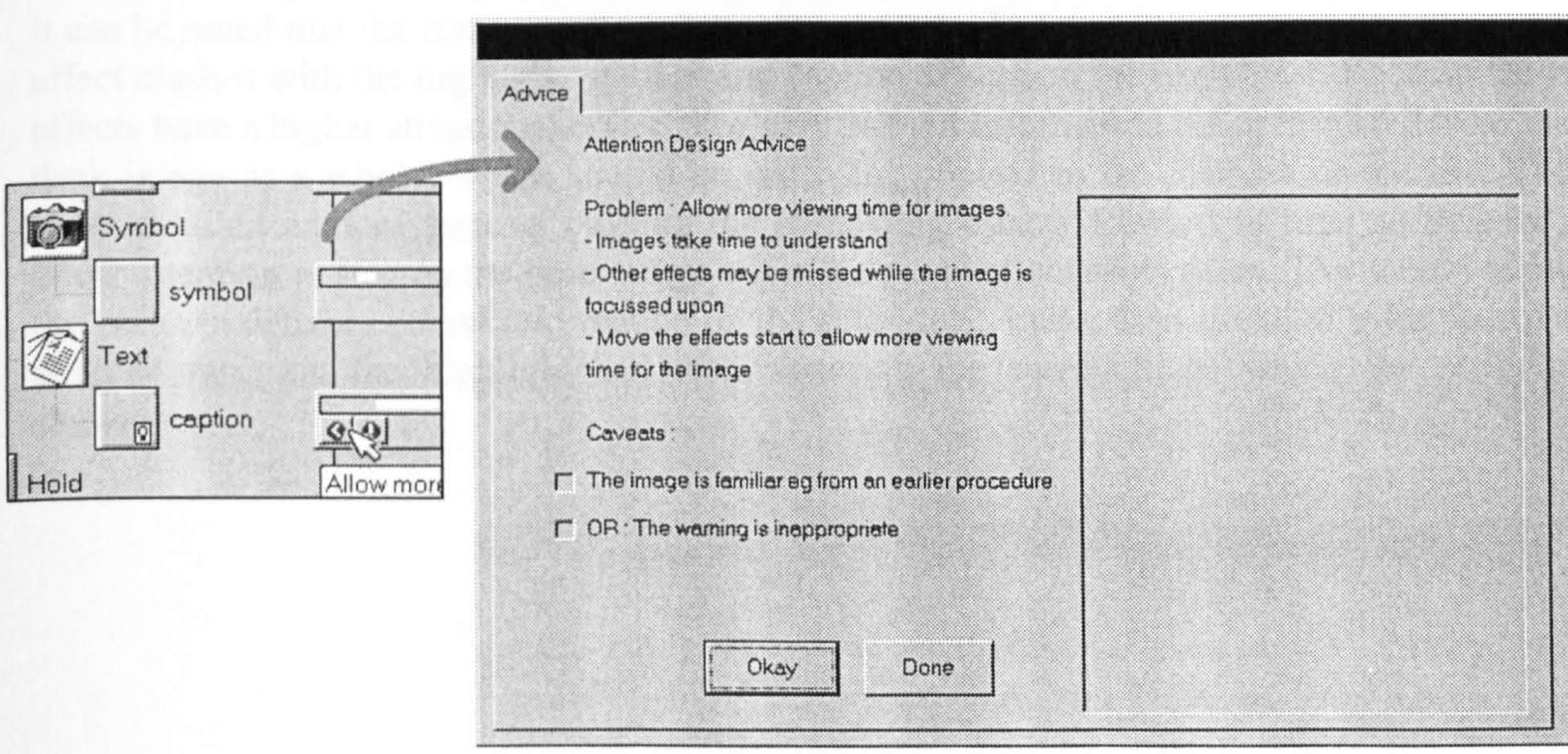


Figure 7.5.2 Detailed design advice dialog

Walking through the attention graph (figure 7.5.1), it can be seen that focus is initially given to the fire image, due to the symbol and highlight effects giving it a higher attentional value than the ship deck image. Focus is held on the image for a small amount of time to allow for viewing.

The rules then try to find other effects which are related to the image, giving focus to the highlight and symbol effects. These have a higher attentional value than the caption, as they are part of the image. The highlight is favoured over the symbol, since it is designed to attract attention, whereas the symbol is a marker. Viewing time is again allowed.

Next focus switches to the caption effect, as it is related by content to the image. The rules then push focus back to allow reading time for the caption. The reading time allowed is based on the length of the caption; it can be seen that the focus bar for the caption is considerably longer than that for the image or symbol. This is because reading takes much longer than viewing images. Following this, focus returns to the image, to form a contact point with the caption. Again, a short viewing time for the image is allowed.

Design advice is delivered by icons, denoting potential problems. The icons provide warnings about problem types, then give more details when clicked. To aid the user, different icon types are used to code warnings as related to effects either prior ('<'), post ('>') or within the timeline bar ('!'). Advice is offered using '?' icons; these do not represent problems, but suggest how to improve the presentation. If the user wishes to disable any of the warnings, this is achieved via the checkboxes on the advice dialogue (see figure 7.5.2).

It can be noted that there are several warnings on the caption in figure 7.5.1. These suggest that the caption should be revealed separately from the image, symbol and highlight so that attention is guided to it in sequence after viewing the image of the fire.

Further warnings are also given concerning the deck object and its image and zoom-in effect. It can be noted that the zoom-in effect does not get focus. The warnings suggest that the zoom effect clashes with the highlight, symbol and caption effects on the fire object. Because these effects have a higher attentional effect, they may prevent information being extracted from the deck image as a whole, which should be providing context to the location of the fire. The more detailed advice suggests pushing the competing effects forward in time so that they allow attention to stay on the hold image while the zoom effect takes place. This would allow the zoom to deliver context information on the ships deck, rather than attention being focused straight away on the highlight and icon, ignoring the surroundings which the zoom is showing.

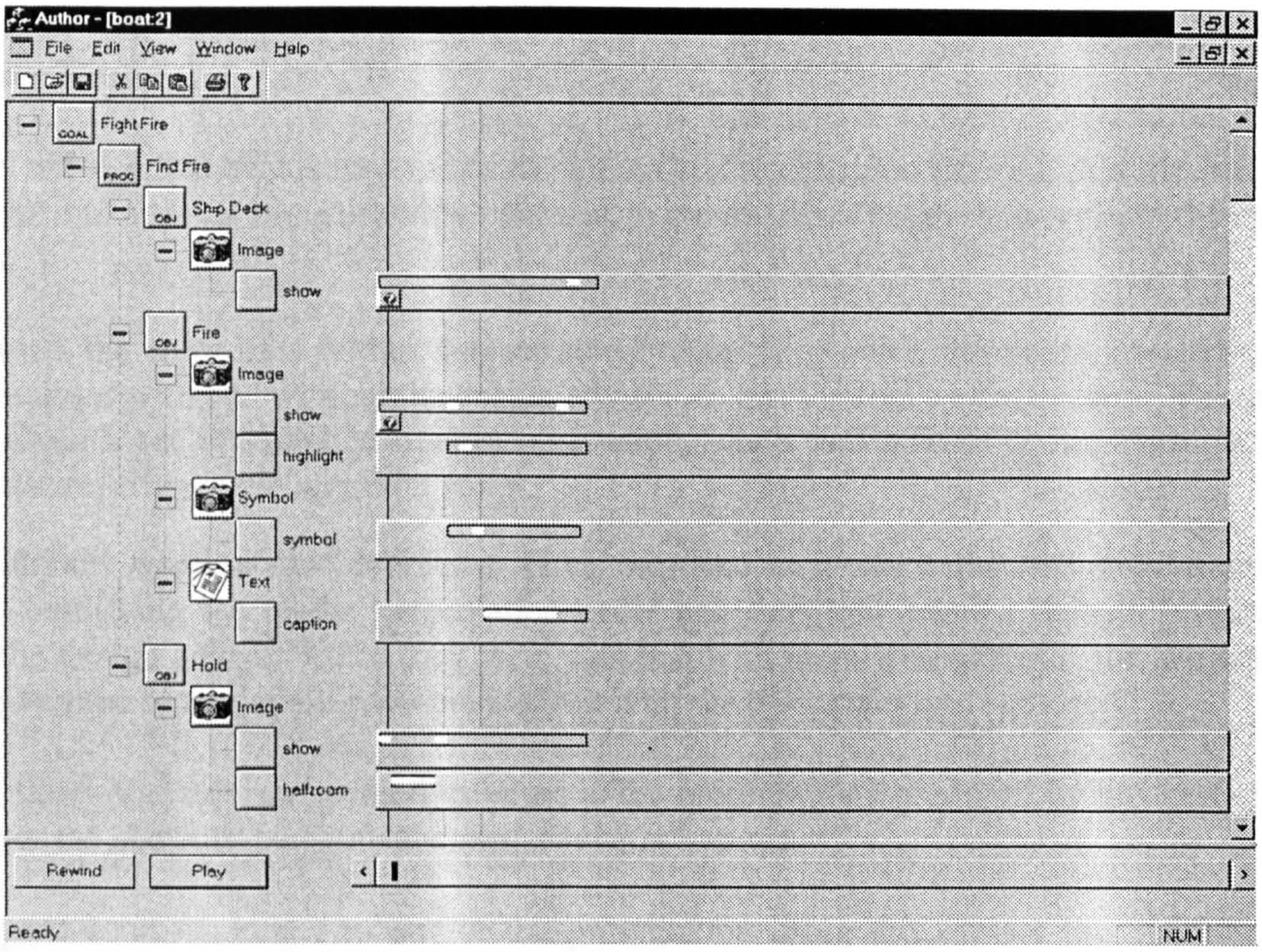


Figure 7.5.3 Re-authored presentation

By adjusting the bars, the problems can be corrected. When all the warning buttons have disappeared, the presentation has no focus problems (figure 7.5.3). An example arrangement is shown below, in which the labels and images are revealed one by one to shift focus around the presentation. Each effect is spaced in time to allow for reading and viewing images. The image of the hold and zoom effect are allowed time to be viewed without the highlight or icon, providing context about the fires location.

This is followed by revealing the highlight and symbol effects for the fire, which will draw attention to the fire object within the image. Finally the caption is revealed and focus switches back to the image to make the contact point. The remaining '?' buttons in figure 7.5.3 warn that images should only be revealed together if they are related to the same topic.

7.5 User Evaluation

So far, only limited evaluation studies have been performed upon the tool. A study was performed to address some usability concerns in the attention graph delivery, and the use of critics.

This took the form of a within subject study using six novice designers, comparing their presentations with and without the design advisor, to test the tool's effectiveness. The subjects were given a set of sketched images, animation, labels and speech for an example cancer presentation arranged in an initial script.

The subjects were set two problems. First, they had to produce the best presentation they could, modifying the media presentation order and durations from a default, without help from the design advisor. Once they were satisfied with their presentations, the subjects were then given the same default presentation, but with the design advisor activated.

Using the design advisor seemed to improve the presentations produced by novice designers, however the subjects' performance raised several issues :

- Subjects took far longer to author the presentation using the advisor than without. They made an average of 30 timeline changes in the passive presentation condition 'v' 120 or more with the advisor version. This suggests that using the advisor makes the subjects take more care and try more design combinations, but does cost in time and effort.

- Subjects suggested that at times the tool overwhelmed them with warnings and advice. The current implementation does not scope advice to a particular part of the presentation, or rate the advice in terms of its diagnostic value. Thus moving one timeline bar can generate a half dozen associated warnings, which may only be connected to one problem. A solution may be to rate warnings by their severity, or cluster them according to their cause. Subjects tended to ignore the more detailed design advice, in favour of the shorter summaries. The detailed advice was used as a last resort, on average only two or three times for the session.

- Two subjects tended to ignore the design advice and used a trial and error approach to removing the design problems. One reported using the tool as 'a game in which the error buttons had to be removed'. This illustrates that design advisors alone may not solve the problem, designers also require background knowledge of the problem to help them interpret advice and motivate their appreciate of why it is useful. Configurability is also important, as two subjects disagreed with one of the advisor's warning, that highlights should be revealed prior to labels, but they were able to disable the warning using the configuration facility.

The study of the tool showed that it did improve design, and that subjects generally found it to be usable. However evaluation of the tool, although demonstrating its effectiveness, did also suggest the need for further studies of how design advice is delivered. The results suggest that designers must also be made aware of the underlying design issues and motivated to use the guidelines, rather than simply trust the tool's advice.

7.6 Conclusions

The current tool is a prototype to allow us to explore how guidelines can be used to assist designers. To this end the tool must present advice which can be readily understood by the designer, and allow the designer to try 'what-if' style exploration. This is supported by a menu driven advisor, which allows the user to easily undo selections, provides advice on design, and can update the task / media model for the designer.

Attentional design is supported by the attention graph. The expert system automatically produces the graph and overlays it on the existing timelines. The attention graph is proposed as a representation which is highly fluid : the designer can simply move the timeline bar and check quickly if any potential attentional problems will arise.

A tool implementation is of value for several reasons. First it allows the method to be used more easily by designers. This should make it an easier task to perform studies on the usability of the method itself, and to get feedback from designers. By embedding design rules and advice within the tool, the designers are no longer required to refer to lengthy paper documentation or to use paper based drawings of task structures and attention graphs. Since the tool is similar to the type of tools which multimedia designers use in day to day life, it should be easier to gain their interest.

Second, in attempting to transform the method guidelines into rules, there is a test in how tractable the guidelines actually are. The guidelines are now required to be explicit about the conditions they would fire under, and the effects that they produce. This is particularly true for the attention graph rules, which required a large number of rules to predict what would be attended to, and to critique problems in the presentation. An implementation gives some idea if the rule set is complete, and that rules do not conflict.

One current concern with the tool is the need to perform further user studies. These would test both how easy its to interact with the design advisor and attention graph, and also ensure that design advice was delivered in such a way as it would be usable to designers. As noted in chapter 6, it is vital to test methods with real world designers. At the current time, only a limited study has been performed.

Chapter 8

Conclusions & Discussion

8.1 Introduction

This thesis proposed a method to improve the design of Multimedia presentations. It was based on a framework which synthesised literature from cognitive and educational psychology and instructional design. Key guidelines were drawn, which were used as the basis for the design method. The validity of the guidelines for improving design of multimedia presentation was tested by studies of eye tracking and comprehension. The usability of the method was then tested with novice designers and formed the basis of an advisor tool.

8.2 Research Contributions

8.2.1 Producing a principled cognitive approach to multimedia design

Chapter one and two of this thesis argue that since the process by which a user comprehends a multimedia presentation is a cognitive one, in order to produce a set of guidelines for the design of such presentations, a framework which integrates perception, attention and comprehension of is required.

The framework seeks to use the cognitive architectures of [Barnard 85], [Baddeley 86] and [Miller & Johnson-Laird 76] to address issues in multimedia design. Thus the features of these models such as attention, short memory, working memory and comprehension are interpreted for their value in design.

The framework has originality in integrating a fragmented literature covering perception and comprehension of a range of media. It was useful in synergising literature from educational, instructional and cognitive psychology into a collection of issues which were relevant to multimedia design. This approach taken in the thesis follows that of other works in HCI such as [Gardiner & Christie 87], who have proposed cognitively based guidelines for UI design. [May & Barnard 95] have suggested how the cognitive architecture of [Barnard 85] can be used to understand how film is designed, and how such techniques can be used for UI design.

The value of using cognitive theories to underpin design is suggested by [Long & Dowell 89]. They contrast 'craft based' guidelines with those from 'applied science' and 'engineering discipline'. It is suggested that a 'craft based' approach is a weak one, in which guidelines are not based on any explicit theory or framework of knowledge, and are thus not generalisable or easily testable. [Long & Dowell 89] propose the advantages of an 'applied science' and 'engineering discipline', in which guidelines are based on explicit and clearly derived knowledge. It is suggested that one source of such knowledge for interface design is cognitive psychology; they note that using such theories provides 'scientific knowledge', which 'is explicit and formal, and so supports the reasoning and derivation of guidelines, their solution and application.'

One lesson that was learnt from trying to build the framework from cognitive psychology was the danger of covering many facets of human cognition without a clear structure of the type of guidelines required. It was important to structure the survey both in terms of the cognitive studies themselves, and also to clearly derive guidelines which would be of benefit for multimedia design. Because the literature is potentially vast, it was very important to have a set of issues around which to structure the survey. These were developed from the architectures of [Barnard 85], [Baddeley 86] and [Miller & Johnson-Laird 76], giving method stages such as attention, working memory and comprehension. Each was then filled out with more specific studies important to multimedia design, such as how attention processes image, text and animation; or how different types of information are comprehended. This analysis lead to a taxonomy of issues for multimedia design, around which the body of the thesis was structured.

Another problem encountered was the need to consider how the framework would cope with the problems of analysing the information to be communicated in a multimedia presentation. Whilst it was clear from chapter one that a content representation was important, it was unclear how to form a typology which could usefully capture the type of content which a multimedia presentation might show. This was in part solved by adapting the work of [Miller & Johnson-Laird 76], who had specified a number of mental model types as way of understanding how language and vision are linked to meaning. These were used as the basis of a set of information types, which formed the core of the framework. Even so, it was necessary to adapt the model types and to limit the type of content which could be analysed by the method to expository presentations.

8.2.2 Providing Guidelines for systematic Multimedia Design

The framework provided valuable insights into where design problems might occur in Multimedia presentations. These are presented in chapter 3 as a set of guidelines which describe the steps to produce an effective multimedia design.

The first part of the method considers how to logically specify the presentation in terms of information types. This provides a considerable advance over other methods of task analysis used in HCI, such as TKS [Johnson et al 1988]. TKS does not offer any consistent way of representing the parts of each task below that of a sub-goal. The thesis is a step forward, allowing the task model to be broken down into information types to provide a more rigorous and fine grained analysis of what constitutes each task step.

The second part of the method concerns media selection, combination and within media design. A key issue here was the use of information types to support multimedia design. The originality here is in the consideration of how several media can be combined to support a particular information type, and how the requirements for communicating the information type is reflected in the decisions taken in 'within media' design.

[Alty et al 92]; [Alty & Bergen 92] establish the principle that different media should be used to communicate different types of content; however, they do not provide such a systematic set of media selection rules, or provide any guidelines as to how to design within the media themselves. For example the guidelines given by [Alty & Bergen 92] suggest that the choice of media leads to the use of text, image or animation, but no detail is given as to the design of these media themselves eg when should a zoom effect be used, or a cut in the animation. The method in this thesis provides guidelines for when each of these techniques should be applied, based on the type of information which is to be communicated.

A problem with the media selection models produced by [Park & Hanaffin 93] is that they do not consider the use of information types to analyse the content to be delivered. Thus [Park & Hanaffin 93] provide highly general guidelines, such as 'organise lesson segments into mutually consistent idea units' and 'present information across multiple, complementary symbols and formats', which say little about how to design the content itself.

Similar problems also exist in [Gagne & Briggs 88]. Again, since they do not have any notion of information type, it is difficult to tell for any given content which media should be used. For example, they suggest that to 'guide thinking', the designer should use 'oral materials','printed text', or 'sound movies'. Without any way of analysing the underlying content [Gagne & Briggs 88] are forced to provide only general advice. The method in this thesis allows an in-depth analysis of the different kinds of information which an animation with sound could deliver against printed text; eg a printed text would be inappropriate if the task was to 'guide thinking' about physical action information.

Worse still is the computational media selection model given by [Arens & Hovy 93]. This acknowledges that the content of the media is important, but state that text, image and animation media all have 'infinite' semantics. This analysis makes it very difficult to understand how a choice of any particular media could be made, or how within media design decisions could be motivated, since it would seem to equate all these media together. An analysis of the content of each media using information types would seem to be of far greater value for design.

Methods which are aimed at evaluating multimedia presentations also suffer from a lack of content analysis. The checklists of [MUMMS 96], [Heller 95] are not motivated by any notion of information types, and thus are largely structured around types of media, rather than the content to be communicated. The checklists do not provide solutions to design problems; instead they simply allows the designer to make a set of judgements. For example [Heller 95] asks the evaluator to rate if the 'use of animation' is 'effective' or 'ineffective'; whilst [MUMMS 96] asks the evaluator to 'agree' or 'disagree' that 'animation was used effectively'. The design method in this thesis provides ways of diagnosing, and solving, such problems by suggesting when animation should be used eg to communicate physical action information.

The 'cook books' approaches of [Hodges & Sasnett 94], [Fisher 96], [Burger 93], [Hoffos et al 91] also have problems in dealing with media selection and within media design. They do not employ any cognitive framework to underpin their design advice; and their guidelines are also not structured around any analysis of content. This leads to guidelines which in some cases are ill-founded. For example [Hodges & Sasnett 94] suggest 'often video serves no purpose other than to offer an attractive display. It is not highly important that the viewer remember the content.', whereas [Burger 93] states 'if a picture is worth a thousand words, a video or animation is certainly worth many more.' Both of these guidelines would cause problems in the design of the presentation, as the method in this thesis suggests the use of animation should be based on the type of content to be communicated eg animation would not be useful for causal information.

In summary, the method aids in understanding of how a combination of media can be used to communicate a particular information type, and how the media can then be designed. This provides a far more detailed level of analysis e.g. how animation communicates physical action information, but may not communicate causal information. The method clarifies when media should be used, and how they will aid the user in extracting the underlying information types.

The third part of the method describes how 'contact points' should be designed between verbal and visual media eg how a speech track should be designed to reference the visuals. Contact points concern how the content of visual and verbal media can be integrated together, based on the type of information to be communicated. Again, the cognitive framework is valuable in suggesting that making links between media may be equally as important as the choice of media itself.

[Narayanan & Hegarty 98] note the need to make referents between 'visual and verbal elements', and 'visual elements with the same referent'. However, they do not provide any account as to how the types of information being communicated can be best combined across the media. For example the method in this thesis suggests that in designing spatial information, the text should provide locations of landmark objects and structure the spatial relations; whilst the image will provide exact locations of each object.

[Andre & Rist 93] give a detailed account the need to design references between text and image, based around a set of rhetorical acts. However, again they do not suggest that particular types of content require particular designs of contact point references. Instead, the provide general guidelines such as 'Elaborate : One part of the document provides further details about another part', which says little about the way the media themselves should be designed. The thesis would suggest that a text could elaborate an animation by providing information concerning causation.

The checklist of [Heller 95] does acknowledge the need for references between visual and verbal information in evaluation questions such as 'how well does the speech relate to the images ?'. However, the checklist does not provide any solutions if the two do not relate, nor does it ask any questions which are targeted at how particular types of content are related which the design method would motivate, eg does the speech track describe the role of each object in the animation.

The design of contact point references is neglected in the models given by [Gagne & Briggs 88], [Reiser & Gagne 82]. They focus on the choice of one or more media, but say little about how the content should be combined. Thus [Reiser & Gagne 82] would suggest using 'sound film' if reading skills are poor and the material requires knowledge of the manner of motion. This provides no advice as to how to design the media together so that the speech track provides the role information for objects in the action eg who does what to who, which is not available from the animation.

[Arens & Hovy 93] also ignore how media can be combined to provided references. In describing the 'semantics' of various media, they posit that text and image both have 'infinite meanings'. However, the method in this thesis predicts that by setting a contact point between the two media, the meanings can be restricted eg if the text describes a particular interpretation of the image, then the image no longer has 'infinite' meanings. This would seem to be a fundamental flaw.

[Rogers & Scaife 98] note 'combinability' between visual and verbal media as an issue for Multimedia design. They suggest that it is important to allow 'users to combine hybrid representations eg enabling animations and commentary to be combined' They suggest 'we could display a diagram with accompanying narration that directs the user to salient parts of the diagram. Such cueing could help the novice also learn how to read the diagram though focusing attention on the relevant aspects'. However, they do not provide any specific guidelines for how to combine different types of information together across media, such as sequence information from text with details of physical actions from the animation.

In summary, the existing approaches either fail to acknowledge the value of designing references between media at all, or do so in an ad-hoc manner. This design method in this thesis provides guidelines as to how to produce contact points to effectively produce such references.

The fourth step of the method is attentional design. This thesis makes an original contribution in suggesting that the rapidly changing nature of an multimedia presentation may pose a significant problem to the serial nature of visual attention. It is argued that for comprehension to be successful, the 'spotlight' of attention must be guided to the relevant parts of the presentation, and care must be taken to avoid overwhelming the viewer. 'Attentional design' provides a set of guidelines which indicate potential techniques to draw the user's attention, and guidelines to sequence the presentation so that attention is able to successfully follow the themes within the content.

Attentional design is only discussed in any detail in two papers : [Alty & Bergen 92] and [Arens & Hovy 93]. [Alty & Bergen 92] studied the delivery of warnings in a process control system. They suggest that a flashing screen or beep can be used as an alert to a particular type of media. However, they do not provide any design techniques for sequencing attention within the medium itself eg the use of a reveal, highlight or zoom in an image to step attention from one part of the topic to the next. Such techniques are vital in guiding attention once it has been initially captured by the medium.

[Arens & Hovy 93] reduce the problem of attentional design to 'detectability', which is either 'high' or 'low' depending on whether attention is likely to be attracted. For example, they specify that animation has 'high' detectability, whilst a text has 'low' detectability. This analysis says little about how attention will be directed in a multimedia presentation, eg if the text is revealed then it is likely to have high detectability, since anything appearing in the visual field will attract attention.

[Rogers & Scaife 98] acknowledge the importance of 'explicitness and visibility' in multimedia design. They ask 'How can we use visual and auditory properties of external representations to make explicit processes that are difficult to grasp because they are not perceptually obvious ?' Unfortunately, they do not provide any specific answers for guiding attention within a presentation. The most detailed suggestions provided are 'flashing arrows could appear at relevant times, to highlight elements of the diagram that are being referred to.' The attentional guidelines in this thesis would caution that whilst animation does gain attention, it also tends to make it difficult to focus upon other the parts of the presentation, which in this case the viewer is supposedly being directed to.

[Narayanan & Hegarty 98] also note the need to consider attention in design, particularly between animation and text. They advise 'if text is provided visually on the screen with animation it presents a situation in which people can either attend to the picture or text, but not to both. Because of the real time nature of animation, a person might miss part of it while reading the text'. However, [Narayanan & Hegarty 98] do not provide any guidelines as to how to design for attention within the media itself. For example, the guidelines in this thesis would note that text is very useful, even in animation, eg if the content of the language is complex; and would suggest that a freeze frame effect is applied to the animation before the text was shown, then reading time allowed for the text, before the remaining part of the animation was shown. This thesis thus provides a finer grained analysis of the design problem.

Presentation planning systems, such as [Feiner 85] , [Feiner & McKeown 91] and [Andre & Rist 93] do use attentional techniques such as highlights and labels. However, these techniques are not sequenced in time, and the rules for the choice of any particular technique are not documented in the published papers. Instead the presentations produced by these tools are simply shown all at once, and the decision of whether to use eg a highlight or arrow effect appear to be ad-hoc.

To conclude, the method provides guidelines for media selection, within media design, and attentional and contact point design which have originality and value in producing a multimedia design that will communicate its content successfully.

8.2.3 Demonstrating that the Guidelines improve multimedia design

The value of the guidelines was empirically tested in chapter four and five. In chapter four, eye tracking studies were used to investigate how attention copes with complex multimedia presentations. Whilst a number of eye tracking studies have been performed on still images [Yarbus 68] and text with images [Hegarty & Just 93], no existing studies had considered how peoples' attention follows a complex multimedia presentation, or postulated that this may give rise to design problems. The eye track data showed that within media design techniques had an effect on what was fixated on, and provided evidence of how design problems could overload attention. These studies are thus of considerable novelty.

In chapter five, studies were conducted upon an existing commercially produced multimedia presentation. By testing user comprehension the effects of design problems were shown in poor recall scores. Because comprehension is related to the user's knowledge, these studies also made use of domain novices and experts. The studies showed that even experts in the domain had problems when recalling the information provided in the presentations.

The guidelines were then used to re-author the presentation. Changes were made in the design of text, speech, images and animations within the presentations. The end result was again tested with domain novices. The results showed that application of the guidelines could greatly increase recall. This provides a validation of the worth of the guidelines for improving design. By re-authoring existing content the studies are also more useful than those performed by [Mayer & Anderson 92], [Large et al 96] which made all or nothing changes from eg still image to animation. These did not distinguish how particular information types in a media are comprehended, or how the media content should be designed to support the type of information being presented.

Further studies were performed using text and speech only presentations. This showed that change in the script of presentation alone was not responsible for improvements, but that careful design and combination of each of the media elements were vital.

Taken together, these empirical studies show that the guidelines could be tested to locate and solve multimedia design problems. By systematically varying the design of a particular media, the studies given in chapter five provide evidence that the guidelines can have marked effects on subject's recall.

8.2.4 Testing the utility of the guidelines with multimedia designers

It is important to test whether designers can make use of the method. Chapter six provides a study of designers which showed that the method was usable, and could help produce better multimedia presentations.

Relatively few attempts have been made to test out the benefits of design methods with their target audience. A study of the use of the ISO standard for menu interfaces by [de Souza & Bevan 90] found that 11% of the guidelines were invalidated, and that designers had difficulty in interpreting 30% of the rules. [Thovtrup & Nielsen 91] performed a study in which designers were given an example interface to diagnose for problems; they found that on average only 4 out of 12 problems present were found. [Thovtrup & Nielsen 91] also note that the ISO guidelines used by [de Souza & Bevan 90] were improved as a result of their study with designers : 'the main lesson from this study is the need for usability testing of usability standards'. For these reasons, the empirical validation of the usability of the guidelines is valuable.

The study showed the designers were able use the method, and reason about multimedia design problems. The method increased their ability to diagnose design problems. If a presentation fails to deliver its content, designers using the method could gain a clearer understanding of the errors by reasoning about which method step is causing problems. Once a problem has been diagnosed, it could be removed without unnecessarily varying other parts of the design.

The results of this study showed that the method can be used by those who were not involved with its derivation. The method was felt to be useful and could be applied without excessive cost. It raised useful issues about the need to tailor method advice differently for novice and expert designers; since experts require more detailed advice, and may ignore rigidly prescriptive methods.

8.2.5 Delivering the Guidelines via an authoring tool

Chapter seven reports the tool design and implementation. This shows that method and its guidelines can be codified in a way that can be operated upon in a general way. The tool also provides added benefits in reducing the cost of using the method, since the designer is no longer required to read lengthy method manuals or use paper based representations. It also makes the method's application more direct by embedding it within an multimedia authoring environment.

Several studies of guideline usability have proposed the value of tool delivery. [Tetzlaff & Schwartz 91] provide a summary of problems with the use of a design guideline manual. They propose that 'to assure conformity, guidelines should be developed in conjunction with the supporting environments and toolkits'. A similar solution is also proposed by [Thovtrup & Nielsen 91], who conclude 'to increase the usability of user interface standards, we recommend having development tools that support implementation of interfaces that follow the standard'. These suggestions point to the value of investigating how guidelines can be placed with the tool itself.

A number of tools have combined guideline advisors with design tools e.g. IDA [Reitner 94], SIERRA [Vandervonckt & Bodart 93]; but only one existing tool, eMMa [Nakakoji et al 97], concerns multimedia design. The advice in the eMMa system is, however, limited to one particular phase of the design, media selection. No advice is targeted at how to combine or sequence the media together.

The tool described in this thesis provides a far wider range of features, such as advice on within media design, attentional design and the design of contact points; and provides a critiquing facility as well as an expert system shell.

8.2.6 Benefits of critiquing for guideline delivery

The design of the tool itself could have followed several approaches : a hypertext manual, an expert system advisor, a planner or critiquing system. Whilst several AI planner systems, COMET [Fiener & McKeown 91], WIP [Andre & Rist 93], and one expert advisor, eMMa [Nakakoji et al 97], exist to produce multimedia presentations; however no current systems have been produced which support critic based authoring of multimedia presentations.

The value of critic systems over hypertext manuals or expert advisors is that the critic system is 'active', so that it can spot problems in the design without intervention from the designer. This is a major benefit in a design environment, in which advice is required continuously as the design is changed. If advice is provided in a 'passive' way, i.e. when the user asks for it, then design problems will build up and their underlying root cause may be obscured. The eMMa system [Nakakoji et al 97] requires user consultation to provide advice on media selection; so if the user adds a range of media to their presentation and seeks advice they may be overwhelmed because the system can only respond to potential problems when the user solicits advice.

There are several reasons for favouring critics over the 'strong' AI approach taken by planners such as COMET [Fiener & McKeown 91] and WIP [Andre & Rist 93]. First, many design problems are ill-defined, and that the complete specification is only available following a process of iterative design. AI based planning systems, which require all inconsistencies to resolved and need a fully defined problem, are unsuitable for many design tasks. Multimedia design is an iterative process which requires that the designer try out different, and not necessarily complete, combinations of media and arrangements of the media over time, none of which may be an ideal solution. Critiquing is therefore useful because it can spot potential problems in the partially defined design as it evolves.

Second, the cost of producing a full specification for a planner to operate upon can in some cases be as expensive as producing the finished design itself, since much of the representation of the task and its users are held by the designer, not the system. In order to capture this information for a planner to make use of it essentially requires the designer to do much of the work before hand. Unlike WIP or COMET, the details of the contents of the media and user do not have to be fully specified for the critic to provide advice.

Third, AI based systems rarely include a notion of the designer; they take a representation of the problem and apply a set of rules to reach some solution state. However, if the solution produced is deemed to be non-optimal by the designer, it can be difficult to understand exactly how the rules applied should be changed. The systems are thus 'opaque' to the designer and leave little opportunity for iterative changes. The 'active' nature of the critics means the designer can try out what-if style scenarios, gaining an understanding of how changes in the design effect the advice. The critic can also be tailored by the designer, allowing further control over the advice delivered.

These concerns appear to be significant problems facing a multimedia designer using systems such as COMET or WIP, suggesting that a critic approach would be useful rather than an expert advisor or planner. Whilst passive tools such as eMMa are a step in the right direction, allowing the designer to be part of the design process; they lack the active ability of a critiquing tool. The implementation of an multimedia critiquer tool thus represents an original contribution, since no other tool provides similar facilities for multimedia development.

8.3 Scope

8.3.1 Scope of Design Method

The method guidelines were based on a synthesis of literature from cognitive and educational psychology and instructional design. This means that there must be some caveats on the accuracy of the advice, given that many studies in these fields are applied to very restricted problems, and few were directly studying the design of multimedia presentations. The literature was selected based on its applicability to multimedia design where possible, and does not represent an exhaustive study of the complexities of the cognitive issues in these areas. For this reason, empirical studies to establish the value of the guidelines in diagnosing problems in multimedia presentations were carried out.

The method is scoped to deal with the presentation design of expository materials such as how-to-do-it style presentations. The justification for this was that task related training would be an important sector for multimedia to exploit. This means that the method does not address several areas :

- Interaction design : the method does not offer guidelines for design of interaction, such as how to link between sequences or script different paths through a sequence.

- Designing content : the method does not address the design of the content itself below the level of information types.

- Emotional or Aesthetic values : the method does not address the design of entertainment titles, or presentations which are designed to have an emotional or aesthetic impact. In many cases, these materials produce their effect by breaking rules, or using media in unconventional methods, which are outside the scope of the method.

By not including interaction design within the scope of the method, instructional strategies such as re-inforcement, simulation and exploratory learning are outside the scope of this thesis. Such strategies are undoubtedly useful for multimedia design; however they all require the ability to produce effective multimedia presentations around which interaction would take place, so the results reported in this thesis provide a vital first step. Effective delivery of information is a necessary building block for any pedagogical strategy.

8.3.2 Scope of Validation studies

The scope of the eye tracking studies were exploratory i.e. they used an existing product and analysed design problems to create the guidelines. The results showed that subject's attention was influenced by the design of the presentation. There are several restrictions on the scope of this work. First, the studies used a relatively small number of subjects. This was because the studies were difficult to perform, and the eye tracking equipment was unavailable for large trials.

Second, the eye tracking studies in chapter 4 did not consider the re-authored presentation, so they are not proof that the design guidelines would fix a particular attentional problem. Such follow-up studies on the re-authored presentations were not undertaken because of the cost in time of analysing the eye track results, and because the re-authoring may result in large design changes e.g. addition of media, which may have led to very different eye tracking patterns and hence obscured the effect of particular attentional design guidelines.

Third, because of the large number of possible combinations of media, presentation effects and sequencing, the eye tracking studies did not attempt to systematically validate each attentional guideline. Instead, the studies used an exploratory approach in which an existing product was used and an attempt was made to detect examples of good and bad design. The results were used to explore how subject's attention responded to the various elements of the presentation. The studies did not attempt to factor out the effects and compare them systematically. This granularity of study would have proved overly time consuming to apply to the large number of design guidelines.

Comprehension testing and re-authoring were used to test if the guidelines could improve a presentation. Improvements in recall provide evidence for the value of the guidelines. The only factor which was controlled was change in the script, which was tested for using a speech and text only condition. Other within media design factors were not tested on an individual basis. Again, the guidelines suggested a large number of potential design changes, which would have proved overly time consuming to study individually. Also, the tests performed upon the re-authored condition used only novice subjects. This was because it was difficult to gain access to further subjects who had a similar background to the original expert group.

8.3.3 Scope of Method Usability studies

The method usability studies may have benefited from a comparison of the presentations which the designers could produce without the help of the method, against those informed by the method. This would have provided a baseline to judge the method's effectiveness in improving design.

The studies also only touched upon the problem with delivering guideline advice to expert designers, rather than novices. The initial results of the method usability study showed that expert designers resented being given method based advice and questioned the value and derivation of the guidelines. The study did not compare how an expert group would cope with producing the presentation against a novice group with and without the guidelines.

8.3.4 Scope of Tool Implementation

The author tool was produced to demonstrate how the method could be used within an authoring environment, and to investigate critiquing and design advisors within multimedia design. As such it does not provide a complete implementation of the method, in particular it covers only a limited number of presentation effects and does not support video or non sketch images.

Embedding the rules within a tool facilitates testing of their effectiveness and completeness. This is particularly true for critiquing, where the predictions of a large number of rules can be tested by using setting up different presentations and analysing the problems the tool suggests. With a large number of guidelines, it is vital to test how they work together, and whether they potentially conflict. The use of the tool as a validation of the method thus shows that the guidelines can be refined to the level of rules and are tractable for implementation.

A central issue is to test the tool's delivery of the guidelines to designers. So far only limited studies have been conducted upon the usability of the tool. These have largely concerned the usability of the critiquing rules with novice multimedia designers.

8.4 Future Work

8.4.1 Design Method

There are several areas in which the design method could be improved in future work. The first issue of interest is what level of information types are most useful for design. There is a potential for adding a range of further types which would provide a more detailed account of how to design the communicative effects of a presentation, such as argumentation. This could make use of existing work, such as found in Rhetorical Structure Theory [Mann & Thompson 88]. However, providing a wider set of information types is likely to increase the burden of applying the method, which may make it less appealing to designers. Another problem with applying RST is that it is orientated towards analysing texts, and so is not easy to apply to images or animations.

Next, there is a need to consider if the method leads to a tendency to 'over design' the presentation i.e. add too many media and presentational techniques. Currently, the method generally favours applying a range of media and presentation techniques. However, if used over-zealously, this could lead to the end user being overloaded with attentional design and contact points, which may distract or confuse. Contact point and attentional design techniques may be best employed where it is vital to make sure that the message is understood, such as task based information provision; or when the domain knowledge of the user is low. For more open ended communication, such as education and learning; or where domain knowledge is higher, these techniques will still be valuable for providing a clearer structure, but may become overly intrusive. The method may thus benefit from further guidelines suggesting what are acceptable usage of presentation techniques, and where a potential danger may exist.

It is also important to also stress that the method was principally proposed for a expository tasks, delivered in an environment without excessive noise, and in which the VDU can be seen at all times. In expanding the method, such scoping advice may be useful to ensure that the guidelines are applicable to the task and domain. If the task type or environment was changed, then the guidelines would require some alteration; e.g. in a noisy environment, the use of speech would be less practicable; for an educational task there would be need for more re-inforcement.

There are two immediate future directions for method research. First, the method will be extended to include advice on the design of interaction in multimedia. This would include guidelines for the compatibility of different modalities of input and output, and how to design interaction sequences, such as branching. Part of this work was undertaken in conjunction with Philips Research.

Second, there would be considerable value in revising the method so that it can be used for evaluation of existing presentations. Evaluation would require that the method steps be reversed; so an existing presentation would be analysed from attentional design and contact points, to media selection and information types. As existing evaluation methods such as [MUMMS 96] and [Heller 95] are checklists which provide no advice as to how to fix the problems they diagnose, the multimedia design community would gain considerably from a more systematic evaluation approach.

8.4.2 Validation Studies

There is considerable scope for further studies on the impact of the method's guidelines. Whilst a problem exists in picking out key guidelines to experimentally validate, by focusing on usability problems in a range of existing products it may be possible to pin-point commonalties, and then home in on underlying problems. As noted before, it will be vital to choose only particular guidelines as there are too many to study individually.

The attentional design guidelines would benefit from further studies of how subject's cope with known problems, such as moving between multiple elements in a presentation when cued by a contact point in speech, and how competing presentation effects gain focus e.g. a highlight shown with animation, or an object revealed while the subject is reading text. Such studies of known 'buggy' presentations could aid in analysing possible design problems.

There is a need to focus on differences between users in the attention studies. The studies of the cancer sequence showed some variation in results between subjects. It would be useful to test presentations which were less strongly driven by the narrative, or to test particular parts of sequences with greater numbers of subjects. This may provide some evidence of whether the results can be generalised to large subject groups, or whether significant individual differences exist. One particular issue may be differences in cognitive styles which subject's employ e.g. between 'verbalisers' and 'visualisers'. Studies could pre-test for any differences in such styles of learning, and then attempt to explore if they effect the subject's utilisation of media.

There is also considerable scope for studying how different media combinations effect comprehension, and to validate the need for within media design. For example, it would be useful to compare the effectiveness of a presentation where the content and presentation media were held static, but which varied presentation effects such as presentation with / without highlights or labels. Such results would be valuable in validating which guidelines were more effective.

8.4.3 Method Usability Studies

The work has shown that when preparing a design method, a key issue must be its acceptability to designers. It is vital to test if designers can understand the guidelines provided, and can reason with the method to improve their presentations. On going work with Philips Research has been useful in exposing the guidelines to designers, and in getting feedback on the concerns of industry.

Our initial studies showed that novice designers did like using the method, and found it possible to use it in designing their presentations. However, one key issue is in producing a method which will be acceptable to professional designers, as well as novices. The study of the method's usability in chapter 6 showed that there were difficulties in delivering method advice to professional designers from creative arts backgrounds who employed a 'just do it' approach.

The results of the study suggested the need to provide more detailed examples and argumentation for the use of the method, to tailor delivery of design knowledge to different groups. If the method is to be targeted at an artistic design community, then further work would be required in ensuring its acceptability.

8.4.4 Tool Implementation

Current work has been addressing usability issues in the tool. In particular, studies have been undertaken which examine how the critiquing system was used by novice designers. This has produced a number of tentative conclusions which suggests that higher level organisation of the tool's advice may be required to allow designers to understand how particular problems may be solved. The study also provided some evidence for the need for greater motivation for the guidelines, so that they could be more accurately applied by the designer. This could take the form of a hypermedia advisor embedded within the tool, which could contain examples of good and bad multimedia product design. This may also aid in motivating a closer examination of the critic advice.

There may be a need to provide higher level templates to support design. These would encompass chunks of possible presentation requirements, such as how to locate an object and then show an action upon it, embodying information types, presentation techniques and default arrangements in time. These design patterns could then be tailored for a specific presentation.

To be truly usable, the tool must also support the ability to author interactive presentations. Work undertaken with Philips Research has added simple branching to the tool. Since the tool is aimed at prototyping, it is unlikely that a more complex scripting language would be required.

In the longer term, it may be possible to provide limited support for automated presentation design, or at least providing the designer with a first cut sequence which does not contain attention problems. However, such a system would go beyond the scope of the current critic based rules, requiring an element of planning.

8.5 Summary

This thesis has proposed and answered important questions about designing and evaluating multimedia presentations. It has provided a cognitive framework for attention to, and comprehension of, an multimedia presentation. A walkthrough method is derived which is developed from the cognitive framework. The validity of the method's guidelines was tested by a range of empirical studies. A tool was produced to test the tractability of the method. Both the method and tool were evaluated with novice designers.

References

[Alty & Bergan 92] Alty, J. & Bergan, M. The design of Multimedia Interfaces for process control. Proc of 5th IFAC IFORS IFIP IEA conference on Man Machine Systems, The Hague, 1992.

[Alty et al 92] Alty, J., Bergan, M., Dolphin, C & Schepens, A. Multimedia and process control : Some initial experimental results. Computers & Graphics, v17, No 3, 205-218

[Anderson 83]. Selecting and developing media for instruction. New York, Van Nostrand Reinhold, 1983

[Andre & Rist 93] Andre, E. & Rist, T. The Design of Illustrated Documents as a planning task. In M. Maybury (Ed.) Intelligent Multimedia Interfaces, pp 94-166, AAAI Press, 1993.

[Arens & Hovy 93] Structure and Rules in Automated Multimedia Presentation Planning. IJCAI-93, pp.1253-1259.

[Arnheim 68] Arnheim, R. Art and Visual Perception. Faber & Faber, 1968.

[Ausubel 62] Ausebel, D. The use of advanced organisers in the learning and retention of meaningful verbal language. Journal of Educational Psychology, 51, 267-272, 1962

[Authorware 96] Macromedia Authorware Users Guide, Macromedia USA, 1996

[Armbuster, Anderson & Meyer 91] Armbuster,B., Anderson,T. & Meyer,J. Improving content area reading using instructional graphics. Reading Research Quarterly, 393-416, 1991

[Baddeley 86] Baddeley, A.D. (1986) Working Memory. Oxford : Oxford University Press, 1986.

[Baddeley & Salame 82] Baddeley, A & Salame, P. Disruption of Short Term Memory by Unattended Speech. Jounral of Verbal Behaviour and Verbal Learning, 21, 130-164, 1982.

[Badget & Sandler 89] Badgett, T. & Sandler, C. Creating Multimedia on your PC. John Wiley & Sons, 1989.

[Barnard 85] Barnard, P. Interacting cognitive subsystems : A Psycholinguistic approach to short term memopry. In Ellis, A. Progress in Psychology of Language, v2 197-258, LEA, London.

[Brandt 54] Brandt, H. The Psychology of Seeing New York : The Philosophical library, 1954

[Baggett 89] Baggett, P. Understanding Visual and Verbal Messages. In Knowledge Acquisition from text and Pictures, Mandl, H. & Levin, J.R Eds, Elsevier Science, 1989, 101,124

[Baggett 84] Baggett, P. 'Role of temporal overlap of visual and auditory material in forming dual media associations, Journal of Educational Psychology, 1984, v76, n3, 408-417.

[Barnard 78] Barnard, P. Interacting Cognitive Subsystems : A Psycholinguistic Approach to STM. In Progress in the Psychology of Language 2, Ed Ellis, A. LEA, 1978

[Betsgen & Costermans 94]. Bestgen, Y. & Costermans, J. Time, Space and Action. Explaining the narrative structure and its linguistic marking. Discourse Processes, 17, 421-446, 1994.

[Biederman 81] Biederman, I. On the semantic of a glance at a scene. In Perceptual Organisation, eds Kubovy, M. & Pomerantz, LEA, 1981.

[Biederman & Ju 88] Biederman, I. & Ju, G. Surface versus edge based determinants of visual recognition. Cognitive Psychology, 20, 38-64, 1988.

[Bieger & Glock 84] Bieger, G R & Glock, M D 'The Information Content of Picture-Text Instructions' Journal of Experimental Education 53 (1984), 68-76

[Boohrer 75] Booher, H R 'Relative Comprehensibility of Pictorial Information and Printed word in Proceduralized Instructions' Human Factors 17, 3 (1975), 266 - 277

[Brewer & Dupree 83] Use of plan Schemata in recognition and recall of goal directed actions, Journal of Experimental Psychology : Learning, Memory & Cognition, 9(1), 1983, 117-129.

[Broadbent & Gathercole 90] Broadbent, D.E. & Gathercole, S.E. The processing of non-target words : semantic or not ? The Quarterly Journal of Experimental Psychology, 1990, 42A(1) 3-37.

[Bryant & Subbiah 94] Bryant, D.J. & Subbiah, I. Subjective landmarks in percpetion and memory for spatial location. Canadian Journal of Experimental Psychology, 1994, 48, 119-139.

[Burch 73] Burch, N. Theory of Film Practice. Saeker & Warburg : London, 1973

[Burger 93] Burger, J. The desktop MM Bible. Addison Wesley, 1993.

[Carello, Rosenblum & Grosofsky 86] Carello, C, Rosneblum, L. & Grosofsky,A. 'Static depiction of movement', Perception, 15 (1986), 41 - 58

[Carpenter & Just 92] Carpenter, P.A. & Just, M. A. Understanding mechanical systems through computer animation. Report No. ONR 92-1, Office of Naval Research, Arlington Virginia, USA, 1992.

[Carroll & Bever 93] Carroll.P. et al Visual analysis of cartoons : A view from the far side. In Eye Movements and Visual Cognition. Ed Rayner, K. Springer verlag, 1993

[Cherry 53] Cherry, E.C. Some experiments on the recognition of speech with one or two ears. Journal of Acoustic Society of America, 25, 975-979, 1953.

[Clark 94] Clark, R. Media Will Never Influence Learning. Educational Technolgy Research & Development, Vol. 42 (2). pages 21-29, 1994

[Davis & Marks 93] Davis,B. & Marks, L. Integrative Multimedia Design. Tutorial presented at INTERCHI '93, Amstermdam, 1993.

[DataMonitor 95] Multimedia Publishing Market Opportunities, Ovum 1995

[de Souza & Bevan 90] Souza,F. & Bevan,N. The Use of Guidelines in Menu Interface Design: Evaluation of a Draft Standard, in Proceedings of IFIP INTERACT'90: Human-Computer Interaction, Detailed Design: Menus, pp. 435-440, 1990.

[Dean & Kulhavy 81] Dean, R. & Kulhavy, R. Influence of Spatial Organisation in Prose Learning. Journal of Educational Psychology, 73(1), 57-64, 1981.

[Deathersage 72] Deathersage,B. Auditory and orther sensory forms of information presentation. In Human Engineering guide to equipment design, Vann Cott, H. & Kinkade, R. Washington : Goverment Printing Office, 1972.

[Dinnel & Glover 88] Dinnel, D. & Glover, J. Advanced Organisers : Encoding Manipulations. Journal Of Educational psychology, 77(5), 514-521, 1988.

[Duchan & Waltzman 92] Duchan, J. & Waltzman, M. Then as an Indicator of deitic discontinuity in adults oral descriptions of films. Journal of Speech & hearing Research, 35, 137-1375,1992.

[Duffy 93] Duffy, S.A. Eye movements and complex comprehension processes. In eye movements and visual cognition. Eds Rayner, K. Springer-Verlag. 462-, 1993

[Durso & Johnson 79] Durso, F.T. & Johnson, M.K. naming and remembering proper and common nouns & pictures. Journal of Experimental Psychology : Human Learning & Memory, v5, n5, 1979.

[Dwyer 67] Dwyer, F.M. The effect of varying the amount of realistic detail in visual illustrations. The Journal of Experimental Education, 1967, 36, 34-42.

[D'Ydellawe et al 93] D'Ydewalle, G. & Gielen, I. Attention allocation with overlapping sound, image and text. In Eye Movements and Visual Cognition, Ed Rayner, K. Springer Verlag. 415-, 1993

[Eriksen & St James 86] Eriksen, C.W. & St James, J.D. Visual attention within and around the field of focal attention : A zoom lens model. Perception & Psychophysics, 1986, 40(4), 225-240.

[Faraday & Sutcliffe 93] Faraday, P.M. & Sutcliffe, A.G. A method for Multimedia Interface Design. In People & Computer VIII (eds Alty, JL & Diaper, D) Cambridge University Press, 1993

[Faraday & Sutcliffe 97] Faraday, P.M. & Sutcliffe, A.G. Designing Effective Multimedia Presentations, ACM CHI 97, 272-279, 1997

[Faraday & Sutcliffe 97a] Faraday, P.M. & Sutcliffe, A.G. Multimedia : Design for the Moment, ACM MM '97 (Seattle, USA), 183-193, 1997

[Faraday & Sutcliffe 98] Providing Advice for Multimedia Designers, ACM CHI '98 (To Appear), 1998

[Ferguson & Hegarty 94] Ferguson, E. & Hegarty, M. Properties of maps constructed from texts. Memory & Cognition, 1994, 22, 455-473

[Feiner 85] Feiner, S. APEX an Experiment in Automated Creation of Pictorial Explanantions, IEE Computer Graphics & Applications, 5(11) 29-37, 1985

[Feiner & McKeown 91] Fiener, S. & McKeown, K. Automating the Generation of Co-ordinated Multimedia Explanations, IEEE Computer, 24(10), 33-41, 1991.

[Fischer et al 90] Fischer, G., Lemke, A. & Morch, A. Using critics to empower users. ACM CHI '90, 337-347, 1990

[Fisher 96] Fischer, S. Multimedia Authoring. AP Professional, 1996

[Focal 76] Focal Encyclopedia of Photography, Vol 1, Focal Press, London, 1976.

[Fleming 87] Fleming, M L 'Designing Pictorial and Verbal Instruction : Some speculative extensions from Research to Practice' The Psychology of Illustration : Basic Research, Vol 2, Springer New York (eds Willow, DM & Houghton, HA), 1987, 136 - 157

[Freyd 83] Representing the dynamics of a static form. Memory & Cognition,11, 342-346, 1983.

[Friedman 79] Friedman, A. Framing Pictures : Role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology : General, 1979, v108, n3, 316-355.

[Friedman & Stevenson 80] Friedman, S. & Stevenson, M. Perception of Movement in Pictures. In The Perception of Pictures, Vol 1 Eds Hagen, M. Academic Press, 1980.

[Gardiner & Christie 87] Gardiner,M. & Christie, B. Applying cognitive psychology to user-interface design. Chichester : Wiley, c1987

[Gagne & Briggs 88] Gagne, R.M. & Briggs, L. Principles of Instructional Design, New York : Holt, Rheinhart & Winston, 1988.

[Gernsbacher 85] Gernsbacher, M.A. Surface information loss in comprehension. Cognitive Psychology, 17, 324-363, 1985.

[Glyn, Britton & Tillman 85] Glynn. S., Britton, B. & Tillman, M. Typographic cues in text : management of the readers attention. In Technology of Text. Ed Jonassen, D. Educational technology Publishing, 1985.

[Gorny 97] Gorny, P. EXPOSE : CHI Counseling for user interface design. In Symposium on Computer Aidded Building of Interactive Applications, INTERACT '97, Australia., 1997

[Grimes 91] Grimes, T 'Mild Auditory-Visual dissonance in TV News may exceed viewer attention' Human Communication Research 18, 2 (1991), 268 - 298

[Hardman et al 93] Hardman, L., Rossum, G & Bulterman, D. Structured Multimedia Authoring. In Proceecding of ACM Multimedia '93, Anaheim California, 283-289, 1993.

[Haring & Fry 79] Haring, M.J. & Fry, M.A Effect of Pictures on children's comprehension of writtent text. Educational Communication & Technology Journal,27(3), 185-190, 1979.

[Hartley 78] Hartley, J. Designing Instructional Text. London : Kogan Page, 1978

[Hegarty & Just 93] Hegerty,M. & Just. Constructing mental models of text and diagrams. Journal of Memory & Language, 32, 717-0742, 32, 1993.

[Heller 95] Heller, R.S. & Martin, C. A Media Taxonomy. IEEE Multimedia, Winter 1995, 36-45.

[Henderson et al 89] Henderson. J.M., Pollatsek, A. & Rayner, K. Covert visual attention and extrafoveal information use during object identification. Perception & Psychophysics, 1989, 45(3), 196-208.

[Hillstrom & Yantis 94] Hillstrom, A.P. & Yantis, S. Visual motion and attentional capture. Perception & Psychophysics, 55(4), (1994), 399-411.

[Hochberg & Brooks 78] Hochberg, J & Brooks, V. The Perception of Motion Pictures. In Handbook of Perception and Human Performance X, 1978

[Hochberg 86] Hochberg, J 'Presentation of motion and space in video and cinematic displays' In Handbook of Perception and Human Performance 1 : Sensory Processes and Perception New York : Wiley, 1986

[Hodges & Sasnett 94] Hodges, M. & Sasnett, R. Multimedia Computing : Case studies from MIT Project Athena. Addison-Wesely, 1994

[Hoffos et al 91] Hoffos, S. & Sharpless,G. The CD-I designers guide. McGraw-Hill, 1991.

[Ianella 94] Ianella, R. HyperSAM - A practical User Interface Guidelines Management System. In Proc of 2nd Annual CHISIG, Queensland, Australia, 1994

[Jackendoff 83] Jackendoff, R Semantics and cognition. Cambridge, Mass. ; London : MIT Press, 1983.

[Johnston & Dark 86] Johnston, W.A. & Dark, V.J. Selective Attention. Annual Review of Psychology, 1986, 37, 43-75.

[Johnson et al 1988] Johnson, P., Johnson, H., Waddington, R. & Shouls, A. (1988) Task-Related Knowledge structures : Analysis. Modelling and Application. In People & Computer IV. (eds Jones D.M & Winder R.). Cambridge University Press.

[Jorg & Horman 78] Jorg, S. & Hormann,H. The influence of general and specific labels on the recognition of labelled and unlabelled parts of pictures. Journal of Verbal Learning & Verbal Behaviour, 17, 445-454, 1978

[Kahneman & Treisman 92] Kahneman, D. & Treisman,A. The Reviewing of Object Files, Cognitive Psychology, 24, 172-219, 1992.

[Kaiser, Proffitt, Whelan & Hech 92] Kaiser, M., Proffitt, D. Whelan, S. & Hecht, H. Influence of Animation on Dynamic Judgements. Journal of Experimental Psychology, Human Perception & Performance, 1992, 18(3), 669-690, 1992.

[Kennedy 82] Kennedy, J M 'Metaphor in Pictures Perception', Perception, 11 (1982), 589 - 605

[Kozma 91] Kozma, R.B Learning with Media. Review of Educational Research. 61(2), 179-211, 1991

[Kintsch & Yarbrough 82] Kintsch, W. & Yarborough, J. Role of Rhetorical Structure in text Comprehension. Journal of Educational Psychology, 74, 6, 828-834, 1982.

[Kirsner, Milech & Stumpfel 86] Kirsner, K., Milech, D. & Stumpfel, V. Word and Picture Identification : Is represenational parsimony possible ? Memory & Cognition, 14(5), 398-408, 1986.

[Koopman & Newtson 87] Koopman,C & Newtson,D. Level of analysis in the perception of ongoing instruction : an exploratory study. Journal of Educational Psychology, 1987, 78(2), 212-223.

[Kraft 86] Kraft, R 'The role of cutting in the evaluation and attention of film' Journal of Experimental Psychology : Learning, Memory & Cognition, 12, 1 (1986), 155-163

[Kraft 87] Kraft, R. (1987) Rules and strategies of visual narratives. Perceptual and Motor skills 64, 3-14.

[La Berge 83] La Berge, D. Spatial Spatial extent of attention to letters and words, Journal of Experimental Psychology : Human Perception & Perfomance, 9, 371-379,1983.

[Large et al 96] Large, A. Beheshti, J. & Breleux, A. Effect of Animation in Enhancing Descriptive and Procedural Texts in a Multimedia Learning Enviroment. Journal of American Society for Information Science. 437-448, 47(6), 1996.

[Lemke & Fischer 90] Lemke, A. & Fischer, G. A co-operative problem solving system for user interface design. In AAAI 90, 479-484, 1990.

[Levie & Lentz 90] Levie, W.H. & Lentz, R. Effects of Text Illustrations : A review of research. ECTJ, vol 30, N4, pages 195-232, 1990.

[Lewis & Walker 89] Lewis, C. & Walker, P. Typographic Influences on reading. Journal of Psychology, 80,241-257,1989

[Loftus & Bell 75] Loftus, G.R. & Bell, S. Two types of information in picture memory. Journal of Experimental Psychology : Human Learning & Memory, 1975, 104, N2, 103-113.

[Loftus & Mackworth 78] Cognitive Determinants of fixation location during picture viewing. Loftus, G.R. & Mackworth, N. Journal of Experimental Psychology : Human Perception & Performance, 1978, v4, 565-572.

[Loftus 83], Loftus G.R. Eye movements in reading - Eye Fizations on Text and Scenes. In Eye Movements in Reading, Perceptual and Language Processes, 1983, AP, 359-376.

[Long & Dowell 89] Long, J. & Dowell, J. Conceptions of the Discipline of HCI: Craft, Applied Science, and Engineering, in Proceedings of the HCI'89 Conference on People and Computers V, Conference Theme Invited Keynote Paper, pp. 9-32, 1989.

[Lowgren & Nordquist] Lowgren, J. & Nordquist,T. Knowledge based evaluation as design support for GUIs. ACM CHI '92, 181-188, 1992.

[Macnamara 91] McNamara, T. The psychology of learning and motivation, v27, 1991, 147-185..

[Macromedia Director 96] Macromedia Director Users Guide, Macromedia USA, 1996

[Madsen 78] Madsen, R. The Impact Of Film. Macmillan Publishing, 1978.

[Mandler & Ritchey 77] Mandler, J. & Ritchey, G. Long Term Memory for Pictures, Journal of Experimental Psychology : Human Learning & Memory, 1977, v3, n4, 386-396.

[Mandler & Parker 76] Mandler, J.M. & Parker, R.E. Memory for complex pictures, Journal of Experimental Psychology : Human Learning & Memory, 1976, v2, n1, 38-48.

[Mann & Thompson 88] Mann, W.C. & Thompson, S.A (1988) Rhetorical Structure Theory : Toward a functional theory of text organisation. Text 8(3), 243 - 281.

[Marcel & Barnard 79] Marcel, T. & Barnard, P. Paragraphs of Pictographs : The use of non-verbal instructions for equipment. In Processing of Visible Language. Ed Kolers, P., Wrolstaid,M. & Bouma, H. Plenum, NY, 1979

[Markus 92] Marcus, A. Graphic Design for Electronic Documents and User Interfaces. ACM Press, 1992.

[May & Barnard 95] May, J. & Barnard, P. Cinematography and Interface Design. Proceedings of IFIP Interact 95, Sweden.

[Maybury 94] Maybury, M. Intelligent Multimedia Interfaces, Tutorial given at CHI'94, Boston USA, 1994

[Mayer & Anderson 91] Mayer, R. Anderson,R. Animation Need Narration : An experimental test of a dual coding hypothesis. Journal of Educational Psychology, 83(4), 484-490, 1991.

[Mayer & Anderson 92] Mayer, R.E. & Anderson, R.B. The instructive Animation : Helping students build connections between words and pictures in multimedia. Journal of Education Psychology, 84(4), 444-452, 1992.

[Michotte 63] Michotte, A.E. The perception of causality. New York, Basic Books, 1963.

[Miller & Johnson-Laird 76] Miller, G. A. & Johnson-Laird, P. Language and perception. Cambridge : Cambridge University Press, 1976

[McConkie & Zola 82] McConkie, G. & Zola, D. Visual Attention During Eye fixations while reading. In Attention and Performance XII, 1982, Ed Colheart, M. LEA.

[Meyer, Brandt & Bluth 80] Meyer, B.J., Brandt, D.M. & Bluth, G. Use of top-level structure in text : Key for reading comprehension of ninth grade students. Reading Research Quarterly, 16, (1980), 72-103.

[Mountford & Gaver 89] Mountford, S.J. & Gaver, W. Talking and listening to computers. In Laurel, B. The Art of human-computer interface design, 319-334, 1989, Addison Wesley.

[MUMMS 96] Measuring Usability of Multimedia Systems. The MUMMS Questionnaire. Human Factors Research Group, University of Cork, Ireland. 1996

[Nakakoji et al 97] Nakakoji, K., Suzuki, K. Ohkura, N. & Aoki, A. A framework to support creativity in multimedia information design. In INTERACT '97, 212-219, 1997.

[Nelson & Loftus 80] Nelson, W.W. & Loftus, G.R. The functional visual field during picture viewing, Journal of Experimental Psychology : Human Learning & Memory, 1980, v6, n4, 391-399.

[Newtson & Enquist 76] Newtson, D. & Engquist, G. The perceptual Organisation of Ongoing Behaviour. Journal of Experimental Social Psychology, 12, 436-450, 1976.

[Newtson & Ridner 79] Newtson, D. & Ridner, R Variations in behaviour perception and ability attribution. Journal of Personality & Social Psychology, 36, 1847 - 1858, 1979.

[Newtson, Rinder, Miller & Lacross 78] Newtson, D., Ridner, R., Miller, R. & LaCross, K. Effects of feature change on behavior segmentation. Journal of Experimental Social Psychology, 14, 379-388, 1978.

[Nickerson 80] Nickerson, R. Speech Understanding and Reading : Some differences and Similarities. In Theoretical Issues in Reading Comprhension. Eds Spiro, R, Bruce, B. & Brewer, W. LEA 1980

[Nielsen 92] Nielsen, J. Usability Engineering. Academic Press, 1992.

[Palmer 75] Palmer, S. Visual Perception & World knowledge. Explorations in Cogntion, 1975, Eds Norman, D. & Rumelhart, D. WH Freeman

[Palmiter et al 91] Plamiter,S., Elkerton, J. & Baggett, P. Animated instruction 'v' procedural instructions for learning procedural tasks. International Journal of Man-Machine Systems, 34, 687-701,1991.

[Palmeri, Goldinger & Pisoni 93] Palmeria, T.J., Goldinger, S.D. & Oisoni. Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology : Learning, Memory & Cognition, 1993 9(2), 309-328.

[Park & Hanaffin 94] Park, I & Hannafin, J. Empirically-Based Guidelines for the Design of Interactive Multimedia. Education Technology Research & Development, 41(3), 63-85, 1994.

[Park & Hopkins 93] Park, O & Hopkins, R 'Instructional conditions for using dynamic visual displays : a review' Instructional Science 21 (1993), 427 - 449

[Patterson 82] Patterson, R.D. Guidelines for auditory warnings systems on civil aircraft. CAA Paper 82017, Civil Aviation Authority, London, 1982.

[Perrot 91] Perrot, D., Sadralodabia, T. & Strybel, T. Aurally aided visual search in the central visual field. Human Factors, 33(4), 389-400, 1991

[Pezdek 77] Pezdek, K. Cross-modality integration of sentences and picture memory. Journal of Experimental Psychology : Human Learning & Memory, 1977, v3, n5, 515-524.

[Pezdek 88] Pezdek, K. Maki, R. et al. Picture memory : Recognizing added and deleted details. Journal of Experimental Psychology : Learning Memory & Cognition, 1988, v14, N3, 468-476.

[Pezdek & Stevens 84] Pezdek, K. & Stevens,E. Childrens memory for audiotry and visual information on television. Developmental Psychology, 20, 212-218, 1984.

[Posner 76] Posner, M 'Visual Dominance: An Information Processing account of its origins and significance' Psychological Review 83 (1976), 157 - 171

[Pressley, Levin & Hope 81] Pressley, M., Levin, J. & Hope, D.J. Do mismatched pictures interfere with children's memory of prose. American Educational Research Association, 1981.

[Price & Humphreys 89] Price, C.J. & Humphreys, G.W. The effects of surface detail on object categorisation and naming. Quartertly Journal of Experimental Psychology, 1989, 41A, 797-828.

[Purnell & Solman 91] Purnell, K. & Solman, R. The Influence of technical illustrations on Students' comprehension in geography. Reading Research Quarterly, 277-299, 1991

[Rayner 81] Rayner, K. Visual Selection in reading, Picture Perception and Visual Search. In Attention and Performance X. Eds Bouma, H. & Bouwhuis, D, 1981

[Rayner & Pollatsek 92] Rayner, K. & Pollatsek, A. Eye movements and Scene perception. Canadian Journal of Psychology, 1992, v46, N3, 342-376.

[Reeves 93] Reeves,B. The Effects of Scene Changes and Semantic Relatedness on Attention to Television. Communication Research, V20, N2, 155-175, 1993.

[Reiber 90] Rieber, L P 'Animation in Computer Based Instruction' Educational Technology Research & Development 38 (1990), 77 - 86

[Reitner 94] Reitner, H. A user interface design assistant approach. In 13th World Computer Congress '94 v2, 180-187, Amsterdam, 1994

[Remeze, Rubin et al 94] Remez, R. Rubin, P. Berns, S. Perdo, J. & Lang, J. On the Perceptual Organisation of Speech. Psychological review, 101(1), 129-196, 1994.

[Resier & Gagne 82] Reiser, R. & Gagne, R. Characteristics of Media Selection Model. Review of Educational Research, 52(4), 499-512, 1982.

[Rubin 80] Rubin, A. A theoretical taxonomy of the differences between oral and written language. In Theoretical Issues in Reading Comprhension. Eds Spiro, R, Bruce, B. & Brewer, W. LEA 1980

[Salomon 79] Salomon, G. Interaction of Media, Cognition and Learning. San Francisco : Jossey Bass, 1979.

[Sinitra 90] Sinitra, G. Convergence of Listening and Reading Processing. Reading Research Quarterly, 115-129, 1990

[Smith & Watkins 72] Smith, J. & Watkins, H. An investigation into some aspects of the illustration of primary school books. Typography Unit, Reading University, 1972.

[Smith & Mosier 84] Smith, S.L., Mosier, J.N. , Design Guidelines for the User Interface Software, Technical Report ESD-TR-84-190 (NTIS No. AD A154 907), U.S. Air Force Electronic Systems Division, Hanscom Air Force Base, Massachusetts, September 1984.

[de Souza & Bevan 90] de Souza, F. & Bevan, N. The Use of Guidelines in Menu Interface Design: Evaluation of a Draft Standard Detailed Design: Menus. Proceedings of IFIP INTERACT'90: Human-Computer Interaction 1990 p.435-440

[Spangenburg 73] Spagenburg, R.W. The motion variable in procedural learning. AV Communication Review, 1973, v21, n4, 419-436.

[Spencer 88] Spencer, K. The Psychology of Educational Technology and Instructional Media. London : Rotledge, 1988.

[Stone & Glock 81] Stone, D. & Glock, M.D. How do yound adults read directions with and without pictures ? Journal of Educational Psychology, 1981, v73, n3, 419-426.

[Sweller, Chandler, Tierney & Cooper 90] Sweller, J. et al Cognitive load as a factor in the structuring of technical material. Journal of Experimental Psychology : General, 1990, v199, n2, 176-192.

[Swezey 91] Swezey, R W 'Effects of Instructional Strategy and Motion Presentation Conditions on the Acquisition and Transfer of Electromechanical Troubleshooting Skill' Human Factors 33,3 (1991), 309 - 323

[Sutcliffe & Faraday 94] Sutcliffe, A.G & Faraday, P.M. Designing Presentation in Multimedia Interfaces (1994), ACM CHI '94 (Boston, USA), 92-98, 1994

[Tepin & Dark 92] Teppin, M.B. & Dark, V. Do abrupt onset peripheral cues attract attention automatically. The Quarterly Jounral of Experimental Psychology, 45A(1), 111-132, 1992.

[Tetzlaff & Schwartz 91] The Use of Guidelines in Interface Design, in ACM CHI91, 329-333.

[Theios & Arnheim 89] Theios, J. & Amrhein, P.C. Theoretical analysis of the cognitive processing of lexical and pictorial stimuli. Psychological Review, 1989, v96, n1, 5-24.

[Thibadeau 86], Thibadeau,R. Artificial Perception of Actions. Cognitive Science, 10, 117-149, 1986.

[Thovtrup & Nielsen 91] Thovtrup, H. & Nielsen, J. Assessing the Usability of a User Interface Standard. CHI '91, 335-341.

[Trabasso & Nickels 92] Trabasso, T. & Nickels, M. The development of Goal plans of action in the narration of a picture story. Discourse Processes, 15, 249 - 275, 1992.

[Treichler 67] Treichler, D.G Are you missing the boat in training aids. Film and AV Communication, 1, 14-16, 1967.

[Treismann 88] Treisman, A. Features and Objects : Fourteenth Bartlett memorial lecture. Quarterly Journal of Experimental Psychology, 1988, 40A (2), 201-237.

[Treisman & Souther 86] Treisman, A. & Souther, J. Illusory words : The role of attention and of top-down constraints in cojoining letters to form words. Journal of Experimental Psychology : Human Perception & Performance, 1986, v12, 3-17.

[Vaughan 94] Vaughan, T. Multimedia : Making it Work. McGraw Hill 1994.

[Venturino & Gagnon 92] Venturino, M. & Gagnon, D. Information tradeoffs in complex stimulus structure : local and global levels in naturalistic scenes. Perception & Psychophysics, 1992, 52(4), 425-436.

[Wadill & McDaniel 92] Waddill,P. & McDaniel,M. Pictorial enhancement of text memory : Limitations imposed by picture type and comprehension skill. Memory & Cognition, 1992, 20(5), 472-482.

[Wagenaar et al 81] Wagenaar, W.A., Varey, C.A. & Hudson, P.T. Do audiovisual aid ? A study of bi-sensory presenation on the recall of information. Attention & Performance X. Eds Bouma, H & Bouwhius, D.G, 379-391, LEA, 1981.

[Willows 79] Willows, D.M. Reading Comprehension of Illustrated and Non-illustrated aspects of Text. Paper presented at Annual Meeting of American Educational Research Association, San Francisco, 1979.

[Yarbus 67] Yarbus, A.L. Eye Movements and Vision. Plenum Press, 1967.

[Vandervonckt & Bodart 93] Vandervonckt, J. & Bodart F. Encapsulating knowledge for intelligent automatic interaction objects selection. In Proceeding INTERCHI '93, Amsterdam, 1993.