A Formal Model for Multimedia Presentation

Katherine Morton

Reproduced from *Multimedia Technologies and Future Applications* (1994), eds. R. I. Damper, W. Hall and J. W. Richards. London: Pentech Press, 191-204.

Introduction

The term ‘multimedia’ currently refers both to a technology and its use. The technology is the seamless integration of text, audio, speech output, video and data, all within a single computer controlled system. Developing interactive multimedia systems is becoming simpler; however, developing *effective* systems will centre around two areas:

- describing in formal terms how to build interactive systems,
- understanding how human users interact with incoming audio and visual information.

The simplest computer interactive presentations consist of static audio and video material such as slides with text, graphics, photographs and pre-recorded sound output. This information is under computer control and can be accessed by a user in either a specified order or limited user choice.

This paper describes a way of making multimedia more effective by incorporating techniques from traditional media such as television and journalism to enhance the presentation. The ‘media’ are concerned with conveying information in an entertaining way and has developed effective techniques for this purpose. In this paper, suggestions are made about how to formalize some of the media techniques. Further suggestions are made about the nature of human audio and visual information processing.

1. The Usefulness of a Formal Notation

The ‘media’ are usually thought of as ‘arts’ based, hence subjective, intuitive and inexplicit. But writing for television and radio is heavily constrained by technology and conventions. In order to import these conventions into a multimedia presentation, it is useful to ask: Are media conventions rule-governed? Because, if the codes and conventions can be externalized and made explicit, and a notational equivalence established between media and formal models, then it should be less difficult to create lively and entertaining presentations by following the rules. Of course, creativity itself is probably not rule-governed, but perhaps the constraints on scripting can usefully be stated as sets of rules.

A first description

The constituents of a media production and a formal statement of those constituents can be seen in Table I.

Formal	Informal
a. defined objects: video clip, audio segment.	a. news reader segments, video clips, news items, interviews with experts.
b. rules characterizing presentation 'movement' depending on object sequencing.	b. storyboard – description of scene sequencing.
c. formal rule-governed sequencing of the presentation, based on explicit hierarchical structure underlying it and showing the relationship between what the 'actors' say, and how they are to be photographed and lit, how the action is to unfold and how they are to move.	c. composite script: • author script (acts, scenes), • camera script (camera shots/angles, lighting), • acting script (action, words), • choreography (movement).

Table I

Writing conventions are clearly stated in prose: writers are instructed as to what type of event can be included in a specific programme, and what constraints are suitable for different genre (types of programme). A set of related explicit statements can in principle be formalized into sets of rules.

> 'Each … represents a kind of special code, shared by its makers and its audience... There are various sets of rules for what should be in a code and how these elements should be used and combined.' Burton, p.76 [1].

The word rule is used here informally, and such rules are not usually fully explicit They are learned by developing writing skills and getting feedback from someone who has internalized the rules successfully. Finally they constitute part of the writer's craft. For the experienced writer lack of precision does not matter, because he or she has internalized the rules and conventions which guarantee consistency. But in the new area of multimedia, interaction of several different media makes the production so complex that explicit rules will be useful to save becoming completely lost. If this is so, then the task is to externalize some of the major rules of media script writing, and to formalize an explicit model of media production to provide a clear framework for the multimedia designer to work with.

For example, narratives have an opening and a closure. This says simply that there is a finite and expressible limit to the story. It is essential to define the time and place of the event, and there are ways of letting the audience know when and where events are occurring, when, where and how they are going to change, and to what or whom the changes are and may be expected to occur. If characters and events are considered in the abstract as formal objects, and if changes occur according to explicit rules established and agreed by both writers and audience, the relations between them could be described by a formal model.

Formalism and Media Scripts

A finite state automaton [2] can be viewed as consisting of:

- a set of states (in which the device may find itself);
- a finite alphabet;
- a function which maps pairs of members of the alphabet and states into the set of states.

The following is an example of how the formalism works:

Suppose the set of states, s0, s1, s2, ... sF, where s0 is the initial state for the device, and sF the final state; suppose also an alphabet, a, b, c, ..., then we can write a rule:

$$d(sM, a) = sN$$

which states that there is a rule describing the behaviour of the device such that when it is in a state sM and inputting (looking at or scanning) an alphabet object a, then it switches state to sN. Figure 1 illustrates this.

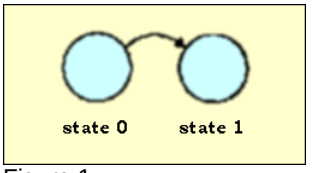

Figure 1

Now suppose our 'device' is a multimedia presentation and our alphabet objects are objects like video clips or slides, then we could say that our device moves between various states with the sequential introduction of various objects.

This type of model can account for essentials such as plot development, dialogue exchange and unfolding information. The entire presentation can be modelled in terms of a progression from state to state according to explicit rules. For example, an educational documentary for multimedia might begin with a statement of a scientific observation. This constitutes the initial state of the presentation — a statement of something as yet unresolved.

A slide-show example

Take the case of giving a talk accompanied by slides. There are two possible approaches:

- slide driven — the slides are presented in sequence, and the speaker talks to the slides
- talk driven — the talk is given, with the occasional slide to illustrate a point.

In the first approach the presentation opens with a slide giving the title for the talk. This is the initial state — state 0; we could say that the presentation is in the title state. The next slide is presented which may show the main points of the talk. This is state 1 — the talk is in the introductory state. The act of putting up the slide (an alphabet object), has by rule changed the talk from its title state into its introductory state. The speaker may now repeat the text shown on the slide, and comment on it. This is an elaboration on state 1, and is represented as a loop off state 1 (see Figure 2). The last comment by the speaker accompanies putting up the next slide, and the talk changes to its next state.

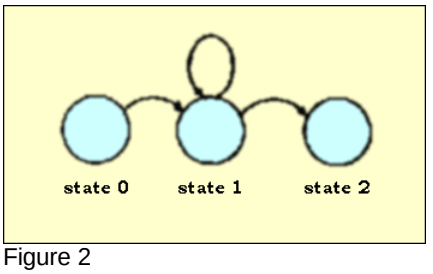

Figure 2

Rules for the slide driven presentation:

- Start with initial state 0, the title with accompanying slide;
- Move to state 1 after 10 seconds by presenting the next slide;
- Use a voiceover to accompany state 1 (loop off state 1);
- Voiceover initializes move to state 2 by presenting the next slide; etc.

In the second approach, talk driven — see Figure 3 — the presentation begins with the speaker stating the title of the talk — the talk is in its initial state. The speaker begins an introduction to the subject, perhaps listing the main points. The talk moves to state 1. A slide is shown, as an accompanying illustration to what is being said and can formally be represented by a loop off state 1. The end of the section on main points initializes state 2. The speaker begins to talk about the first main point of the talk.

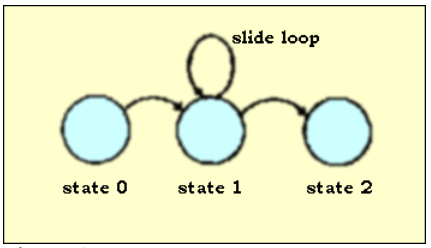

Figure 3

It is possible to change the sequence slightly; the slide could be shown accompanying both state 1 and state 2. That is, remain on the projector for some time because this is a description of the states for the talk, not of the slide. In that case, a loop off state 2 will represent the slide accompanying the talk in state 2.

Rules for the talk driven presentation:

- Start with initial state 0, speaker tells us his title;
- Move to state 1 by telling us about the talk's main points;
- Show slide listing these points (loop off state 1);
- Move to state 2 by starting to talk about the first point; etc.

This illustrates how the elements of the talk — speaker and slide — are in a different relation to each other depending on which one is seen as the 'alphabet' that the talk is scanning. That relationship can be described by the rule generalization:

d(talk state, next slide) = next talk state

d(talk state, next spoken topic) = next talk state

A documentary example

A more complex example: take a teaching documentary on the coastline erosion in East Anglia. The presentation opens with a video clip of sea pounding cliffs, and captures an actual cliff fall. This is the initial state of the presentation — state 0 [sea pounding cliffs]. Partway through the video clip a voiceover suggests the problem is general along the coastline; this voiceover object triggers a transition to state 1 (widespread coastal erosion in East Anglia) with stills to show several examples of eroding coast line, confirming the problem is widespread. State 0 has completed a transition to state 1 — a full statement of the subject of the media presentation. That transition has been triggered by shifting the type of scene in the video clip from particular to general, and by the voiceover suggesting that state 0 is an example of a general situation to be illustrated in state 1 of the programme. In this case the voiceover is an alphabet object.

This can be expressed formally: if the presentation is at state 0 and an influence external to that state (the voiceover) impinges on that state, then a new state will arise — state 1. Thus

Rule: state 0, encountering an external influence, I, changes to state 1

As before, the rule can be generalised as:

d(sN,I) = s(N+1)

which reads as: there is a rule such that a production in a state n, subject to an influence I, changes into a state n+1. sN and s(N+1) are members of a well-defined set of states and I is a

member of a well-defined set of influences, or alphabet. State 0 is singled out as the initial state of the device and state F as the final state in which the production finds itself at the end of all influences.

This formalism views the media production as a device (in the formal sense) existing in a given state, which scans an input which prompts a switch to another state. Only inputs from the defined Set will cause a switch in states, and the only state changes possible involve those which are available in the set of all possible states for the device and which follow the set of rules.

This is an example of a particular rule for the start of an educational media presentation. The presentation will have many states and many rules — all conforming to the general case (see Figure 4).

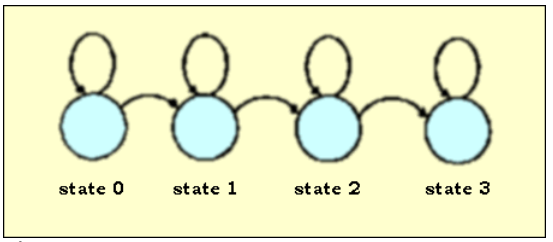

Figure 4

The next stage in the production could involve a brief interview with someone who has lost property to erosion. A voiceover towards the end of the time assigned to state 1 influences a transition to state 2 — the interview clip. While the interviewee is still talking, the shot of him dissolves (the influence) to more pounding waves as the interview ends dissolves to another interviewee as the voiceover introduces an expert on the subject — state 3. The expert will confirm that erosion as presented in states 0 and 1 are possible, and that state 2 is widespread.

Rules:

- Start with an initial state, making a clear point.
- Use a voiceover to make a transition to a first state showing the extent of the point being made.
- Use a voiceover to introduce0 a second state — reconfirmation of the point being made, but from another medium (video evidence moves to human victim statement using language).
- Use a voiceover, together with video clip echoing the initial state, to move to a further state — same medium (the interview), but this time with a detached speaker (the expert) confirming the previous state (interview with victim) and the opening states (erosion is a problem).

A news example

The model so far describes events occurring in sequence at one level only: but it is useful to capture generalizations and talk about these generalizations at different levels. Generalizations are related in a hierarchy — the most abstract down to particular events — the actual events that occur in the world. A tree structure describes these relations.

The formal approach can be applied at any level in the hierarchy; the formalisms will apply at each level. For example, Level 3 may be described by the following (see Figure 5).

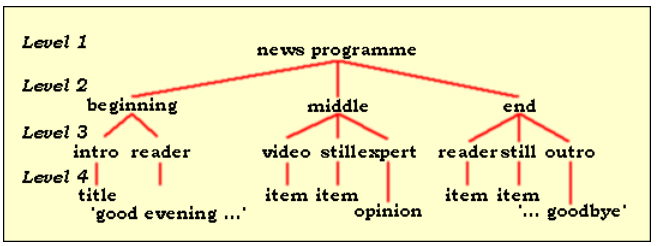

Figure 5

State 0 is the initial state, the title of the news program. Note that the actual title as it appears in Level 4 can be different each time. Level 3 is concerned with the idea of a title, an abstract title. The title may also include sound.

A transition to state 1 results in a video of a news reader. A final statement from the reader initializes state 2. As with the tide, the idea of a formal statement will be there on each occasion, although the statement itself will vary. State 2 is a video of the event reported in state 1; a transition to state 3 of a still, accompanied by voiceover, initializes state 4. The transition to expert opinion on the event is state 3 may be accompanied by another opinion or comment 'This is a loop off state 4'. A transition to state 5 brings a reader back to screen and a new news item. A final statement by the reader initializes state 6, a still, and a voiceover initializes state 7, the exit or 'outro' of the program. This itself can introduce another program, such as weather, or ads, which will itself be initially in state 0.

The description here is of a level in the program, not the state sequence of the reader, the video or the cameras. Each of these may have their own state model equivalent to the appropriate scripts as written for media productions.

Rules:

- Start with the tide of the program.
- The news reader reports on the main news.
- A video is shown of the main news event.
- Voiceover triggers a still of the event.
- Voiceover triggers expert opinion.
- Another comment may be made by some one involved.
- Back to expert opinion.
- The news reader reports on a second item.
- Voiceover triggers a still.
- News reader's outro signal end to program.

The top level, Level 1, defines the genre, the type of program, and requires a set of metarules to distinguish it from other genre.

2. Genre

When considering media, several types of programme are defined, called 'genres'. There are about six major categories of genre: news, documentary, soap opera, quiz, drama, and some types of program that fall into two categories e.g. docudrama [3]. Various descriptive terms can be assigned to define the genres: such as repetition, circularity, openendedness, etc. Restated in formal terms, these descriptions are seen as metarules — general conditions which define genre in the same way metarules define flora and fauna in biology.

Descriptive terms

There are a number of these terms; the following are a selection which can be used informally for defining useful genre types for multimedia.

Repetition The elements that define a particular show are repeated and reinforced during the time the show is running. Formats are repetitive; for example, a newscast covers what is happening in the world at the moment, a police drama always covers crime.

Novelty Accompanies repetition, otherwise the viewer becomes bored. A new problem must be presented each time; e.g. in news, there must be new events or disasters, human interest, etc.

Openendedness Unresolved problems with future solutions.

Immediacy Gives the audience the sense of being in the action, of finding a resolution to a problem presented earlier.

Segmentation Multiple events, multiple stories which weave together as in a drama. Also refers to different types of media objects — video clips, stills, talking heads, etc.

Metarules

The above relatively informal descriptions of the occurrence of particular parameters can be reformulated as rules for defining particular genres (Level 1 of Figure 5):

- If news, then immediacy **AND** openendedness **AND** repetition **AND** segmentation **AND** novelty.
- If drama, then immediacy **AND** repetition **AND** segmentation **AND** (optionally) openendedness.
- If documentary, then segmentation **AND** (optionally) openendedness **AND** (optionally) repetition.

Subrules

- if immediacy, then include alphabet objects to enable audience involvement.
- if openendedness, then final state to include (at least one) unresolved issue.
- if repetition, then a 'feature' to re-occur (e.g. format, transmission time, etc.).
- if segmentation, then 'story' to be multi-stranded.
- if novelty, then unexpected elements to be included.
- Result of defining genre by applications of the rules (see Table 2):

Description / Genre	**News**	**Drama**	**Documentary**
Immediacy	yes	yes	no — only some are about current issues
Openended	yes	not usually — unless series	can be — if about current issues and ongoing problems
Repetition	yes — predictable format — short items, news flashes, main items, weather and sport	characters have goals, change, and learn; types have costumes	some are — formats different, although can expect resolution to problems or statement about future
Segmentation	yes — studio, stills, video clips, expert opinion	usually more than one story	yes — clips, narrator —all feeding into main theme — expert opinion
Novelty	yes — after time	yes — new twist on old themes	can be — underwater photography, unusual topics and investigators

Table 2 — Genre

3. Perception of Audio and Visual Presentation

Media techniques affect the viewer and listener emotionally: they arouse happiness, anxiety, envy, sympathy, amusement, dislike etc. Programmes reinforce our view of the world. But television, film, and radio are not interactive — multimedia adds the dimension of interaction with the material presented on the screen.

For a television viewer, incoming information is bimodal. The two sensory modalities are hearing and vision; these modalities mediate the sound and vision information. Sound images are first detected by the ears and vision images by the eyes. They are modelled as passive activities at the periphery. In hearing, a spectral analysis of frequency and amplitude are made automatically in the cochlea. The results of the initial analysis are digitised and transmitted to the auditory cortex.

In vision, the visual scene is analyzed in terms of the coordinate distribution of the amplitude of light falling on the retina. Special receptors on the retinas analyze the frequency (colour) context of the light. The digitised information is transmitted to the visual cortex. The viewer/listener cannot alter this processing — it can only be stopped by shutting the ears or closing the eyes and preventing signals from reaching the transducers (cochlea and retina).

At the next stage, the information is interpreted according to stored information in terms of the individual's previous experience. This experience affects the perception of what is heard and seen.

In modelling perception for multimedia, it may be useful to distinguish between two types of signal in both modalities; signals that encode language and signals which are not language. Current models of perception differentiate between speech and non-speech auditory signals (see Figure 6).

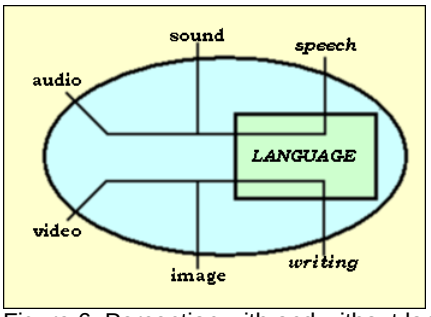

Figure 6. Perception with and without language.

Recent work [5] based on work begun in the early 80's [4] suggests that language information in text and speech is processed differently from non-language sound and vision. There is some evidence to show that humans perceive audio information through two parallel systems; one for speech signals and the other for non-speech acoustic events. These two systems may function at the same time without interference and recognized by the listener as occurring in real time [5].

A multimedia user may interaction through more than two channels. Therefore, in multimedia systems we need to determine if voiceovers, interviews, or teaching are decoded differently from illustrative sound or sound background; and whether written language as in visual text is treated differently by perception processes than graphical or screen images.

We may need to take into account four input channels — sound as in music, speech as in talking, text as in writing, and visual images in video and graphics, (see Figure 6) The strategies for dealing with information presentation may well be different. A formalism would need to take account of these four channels as separate modalities able to combine to influence one another.

The multimedia user has only a two-channel communication system for actively providing information to the system: the mouse, or keyboard. That may soon be augmented by speech recognition systems.

4. Conclusion

The future of multimedia developments is unclear. At the moment there is a great deal of enthusiasm and the future is unbounded. Trying to establish a formal basis for externalizing effective media communication and a formal framework may well help. In this paper, suggestions have been made for explicit rule sets when building a system to transfer techniques from media to multimedia. The rule set might consist of: (1) metarules and subrules; and (2) finite state rules about changes at different levels. Suggestions have also been made that applications models for perception of language, sound and vision could also contribute to effective interactive multimedia presentations.

References

[1] Burton G., 'More than Meets the Eye', Arnold: London (1991)

[2] Hopcroft J.E. and Ullman J.D., 'Formal Languages and Their Relation to Automata', Addison-Wesley: London (1969)

[3] Fiske J. and Hartley J., 'Reading Television', Routledge: London (1990)

[4] Fodor, J., 'The Modularity of Mind'. MIT Press: Massachusetts (1983)

[5] Liberman A. and Mattingly I.G., 'Modularity and the Effects of Experience', Haskins Laboratories Status Report on Speech Research 105/106: New Haven (1991)