CINEMATOGRAPHY AND INTERFACE DESIGN

Jon May and Phil Barnard

MRC-APU, 15 Chaucer Road, Cambridge, UK—CB2 2EF
{jon.may, philip.barnard}@mrc-apu.cam.ac.uk; http://www.mrc-apu.cam.ac.uk/

KEYWORDS: design support, cognitive models, user modelling, cinematography, interface design, multimedia.

ABSTRACT: Interface designers are increasingly relying on craft based approaches to compensate for a perceived lack of relevant theory. One such source is cinematography, where film-makers succeed in helping viewers follow the narrative across cuts which change the information on the screen. Cinematography has evolved over the last century, and its rules of thumb cannot be applied directly to interface design. We analyse film-makers' techniques with a cognitive theory (ICS) and show that they work by preserving thematic continuity across cuts. Expressing this theoretically allows us to extrapolate away from film, applying it to screen changes in interface design.

1. INTRODUCTION

Software designers are being encouraged to produce multimedia applications to make full use of the capabilities of desktop computers. The trend towards more presentationally complex interfaces poses a problem for HCI, since the difference between multimedia designs and the 'desktop' metaphor is vast. Designers are increasingly relying on a craft based approach to compensate for a perceived lack of relevant or applicable theory. An interface technique will be adopted because the designer has seen it, or something similar, used elsewhere. This imitative design does not limit itself to using other interfaces as source material. For example, Koons et al (1992) reported that "as designers, we selected elements and styles from each of the areas from which multimedia is evolving: print, television and computers." Recommendations have been made that designers look to domains ranging from 'form-giving' (Smets, Overbeeke and Gaver, 1994), to Stage Magic (Tognazzini, 1993), to Cinematography (Young & Clanton, 1993). While many of these domains have interesting parallels with HCI, it is sometimes difficult to see exactly how their craft skills can be transferred to the design of interfaces.

In this paper we examine one of these domains, cinematography, and show how the expertise of film-makers can be made relevant to interface design. Our approach is to explain the 'good practices' of film-makers in terms of cognitive theory, to understand why they make a film easy to follow. Our analysis suggests that many film editing techniques have the effect of maintaining the thematic continuity of structural and semantic interpretations. We then extrapolate from this theoretical explanation to interface design, and show that the implications for multimedia contradict some currently fashionable ideas.

2. CINEMATOGRAPHY

Cinematography has surface appeal to interface designers because of the similarity in appearance between interfaces and film. Over the last century cinematographers have developed ways of directing and maintaining viewers' interest in and comprehension of their material. Computer interfaces are frequently perceived as baffling and busy, with too much going on, and relevant information too hard to find, and yet film-makers can cut rapidly between different viewpoints and different scenes without confusing their viewers.

The whole purpose of cutting film is to move the viewer rapidly through a narrative sequence, without waiting for the camera to pan from side to side within a scene, or to move from place to place to follow the action, and to avoid longeurs while nothing dramatic is happening. An edited film presents views that are not spatially or temporally connected. The user of a computer is faced with a similar task to the viewer of a film, in that they must form an understanding of what is happening by observing events displayed upon a screen. The interface designer's task is to build into the computer system the rules that the film-maker uses heuristically to construct comprehensible scenes.

Unfortunately for interface designers, cinematography has developed as a craft skill, rather than theoretically (c.f. Long, 1989). Examination of any number of books on 'film theory' (e.g., Adams,

K. Nordby et al. (eds.), *Human-Computer Interaction*

1977; Bernstein, 1988; Boorstin, 1991) reveals unsorted 'dos and don'ts', listed anecdotally with plenty of examples, but with no real explanation of why certain practices are 'filmic', and why others are 'unfilmic'. There is no way to extrapolate directly from these heuristics, grounded in the medium of film, to interface design, where there are many different constraints. If we were to follow the path of communicating craft skills to interface designers, we would in effect have to teach them how to become film directors, editors and camera crew.

To understand how film-editing techniques can help people comprehend interfaces, we first have to understand how they help viewers comprehend films. As early as 1916, Münsterberg (trans. 1970) compared the close-up shot to perceptual attention; flashbacks to acts of memory and mental imagery; and the sequencing of shots to the sequential direction of attention around a real-world visual scene. Carroll reports (1980, p.20) that another early film theorist, Pudovkin (1929 [1958]) claimed that the role of the editor of a film is to guide the viewers' attention to particular elements of the scene, and that the laws of editing are thus the same as those governing 'ordinary looking'. He and other analysts (e.g., Balázs, 1945 [1970]; Eisenstein, 1949) also discuss the use of close-up shots to magnify critical details to the exclusion of the surrounding scene, in the same way that a viewer in the real-world can concentrate on one part of their visual scene to the exclusion of the periphery of their gaze.

The idea that film construction mirrors the cognitive processing of a viewer interacting in the everyday world is our starting point for the theoretical analysis of cinematography. We have been using a resource based model of cognition called Interacting Cognitive Subsystems, or ICS (Barnard 1985; Barnard & May, 1993). This describes the flow of information through the human cognitive system, from sensation and perception, through central interpretive processes, to action.

3. AN OVERVIEW OF ICS

In general terms, ICS models cognition as a flow of information through different levels of mental representation, each level being transformed to another by independent subsystems. Sensory information in Visual and Acoustic codes is structurally interpreted to produce Object and Morphonolexical representations, respectively. These are in turn transformed by their respective subsystems into Propositional representations that describe the relationships between the elements of the structural codes. This level of representation can feed back to the structural subsystems, providing 'top-down' information to be blended with the 'bottom-up' sensory information. It also produces a higher level, schematic meaning of the material – an Implicational representation. This then feeds back a Propositional representation to be blended with the output of the structural subsystems. There are consequently several cyclical flows of information within the overall cognitive system (Figure 1). These contribute to the stability of cognitive representations, for at any moment the structural representations of the sensory world (whether acoustic or visually derived) are influenced by the currently active Propositional representation, which is in turn influenced by the currently active Implicational representation. In addition, there are direct routes from the sensory subsystems to schematic Implicational representations (not shown in Figure 1), so that the Propositional output of the Implicational subsystem can be influenced by the qualitative nature of the Visual and Acoustic information, as well as the individual's 'body state' (see Teasdale & Barnard, 1993).

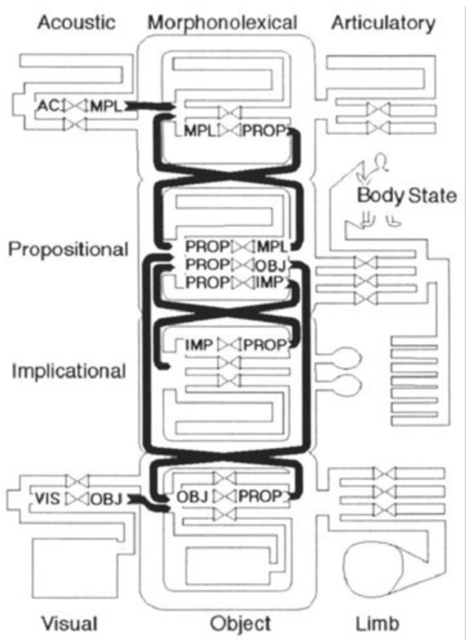

Figure 1: the flow of information active during the comprehension of a film

Each cognitive subsystem is independent, and functionally they are similar, in that they all receive inputs, store and revive inputs from memory, and transform them into other representational forms. This similarity allows us to model the structural details of each representation independently of their level. In particular, we can distinguish between the 'psychological subject' of a representation and the other elements, which form the 'predicate structure'. At the propositional level the subject will be the semantic topic, with the predicate structure containing information 'about' the subject (c.f. Halliday, 1970). For example, the sentence 'John hit the table' has John as its psychological subject, while 'The table was hit by John' has the table as its psychological subject. Both sentences contain the same 'information', but propositionally have different subjects, and so will differ in their interpretation. Similar distinctions can be drawn for the structural levels of representation, where the current subject corresponds to the 'attentional focus'. In an Object representation, this would be the visual object that is being attended to (and so is often spatially constrained), while in the Morphonolexical representation, it would be the part of the sound that is being interpreted (and so is often temporally constrained). This generic concept of structure has been used to analyse visual search tasks (May, Tweedie & Barnard, 1993) and the learning of task sequences (May, Barnard & Blandford, 1993).

The identification of structure in mental representations helps explain how cyclical flows of cognition operate. An architectural constraint of ICS is that any process that transforms information from one representation to another (for example, from an object to a propositional representation) can only operate on one stream of information at a time. For example, it is impossible for the Object subsystem simultaneously to produce two different propositional representations, one based on input arriving from the visual subsystem, and one from input arriving from the propositional subsystem (see figure 1). One input must be 'ignored', or they must be blended into a single, coherent representation – in which case the resulting output back to the propositional subsystem will reflect the conjoint, blended representation rather than the two original inputs (Barnard & May, 1995). To be blended, both of the input representations must have matching subjects.

Representational flow and structure allow us to analyse cognition at a high level, without needing to consider the detailed content of the representations. As such, ICS is well suited to the analysis of general instances of tasks such as film-editing.

4. AN ICS ANALYSIS OF FILM EDITING

In watching a film, the viewer has to interpret visual and auditory information (producing structural representations), and to blend these two streams of information together to construct an understanding of a continuing narrative (propositionally and implicationally). For much of the time this is not problematic, since the two streams are highly correlated, and from moment to moment objects move in a predictable manner around the screen, and sounds happen in tandem with events. In ICS terms, the propositional representations abstracted from the structural representations can be blended together, and they can also be blended with the output of the implicational subsystem.

Potential difficulties occur when a within-scene cut occurs from one shot to another with a different camera position, or when there is a cut from one scene to another, with the visual scene changing completely. Film-makers regard a noticeable within-scene cut as a bad cut, since it destroys the viewers' illusion of 'being there'. By making them aware of the medium of the film, rather than of the scene that is being portrayed, it interrupts their comprehension of the narrative. The skill of the film-maker lies in constructing cuts so that they are not noticed.

Although by definition a cut changes the visual information present on the screen, it is possible to provide a degree of continuity. One technique involves placing the 'subject' of the new shot in the same place on the screen as the 'subject' of the previous shot ('collocation'). Much of the time this allows the camera to move, and close in or pull out, while portraying the same subject. To achieve this, the film-maker will attempt to direct the viewer's attention to the appropriate object before a cut occurs, for example by having an actor look towards it or reach for it. Conversely, where the successive objects must not be related structurally, but should be related propositionally, the film-maker can direct the viewers gaze away from one subject towards the area where the new subject will appear, following the cut. In a conversation between two actors, for example, cuts are not made between identical shots of their faces, but one actor will look across the screen, and the next actor's face will appear there, looking back towards the place that the last actor had been. This avoids the potential confusion of trying to interpret the successive subjects as views of the same object.

The concepts of psychological subject and predicate provide a way of analysing these structural changes. Shifts in visual attention correspond to thematic transitions in the object representation that make an element of the predicate structure the new subject (the previous subject correspondingly becoming part of the predicate). Within a shot, these transitions are under the control of the viewer, and so will be driven by their propositional understanding of the narrative. When a cut occurs, in contrast, the new subject of the object representation will generally be the element of the shot that is nearest the screen position of the previous subject. If this is unrelated to the topic of their propositional representation, they will not be able to relate the two scenes.

Figure 2 shows how filmic and unfilmic versions of a cut result in object representations that are compatible and incompatible with the propositional

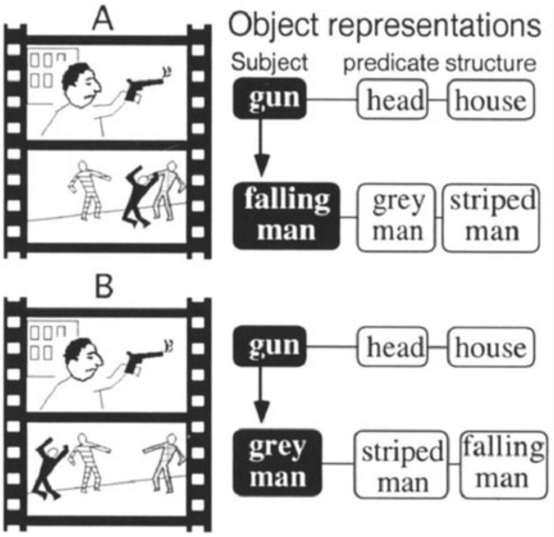

Figure 2: a filmic cut (A) and an unfilmic cut (B), with their object representations (spatial relationships are not represented in this figure).

representation. In version A, the collocated subject fits with the propositional representation ('The Actor is shooting someone', 'Falling Man has been shot'); in version B it fits less well ('The Actor is shooting someone', 'Grey Man looks surprised'), and the viewer must search for a new subject.

The predicate structure also influences the propositional interpretation that is derived from the object representation. A view in a film of a person sitting in a moving car will have a predicate structure that includes elements of the outside world moving past, and this gives some information about the direction that the subject is moving in, even though they may be in a fixed position on screen. Interestingly, this poses problems for film-makers when they have to portray *two* people sitting side by side in a car and holding a conversation, since shots that cut from one to the other will contain predicate structures whose elements move in opposite directions. While this will not be explicitly noticed by viewers, the resulting scene is propositionally incongruous and will appear 'odd'. In practice, cuts are avoided by mounting a camera outside the car window and filming both actors in the same shot.

The general case that can be made from this analysis is that in cutting together two shots it is crucial to provide the viewer with structural representations that can be integrated with their propositional representation. There must be some continuity of the theme across the cut. This is not restricted to the visually derived representations, but also applies to the acoustically derived representation. Suppose that a scene shows a person walking down a corridor. At this point the acoustic information (regular wave-forms with a sharp onset, each peak alternating slightly in pitch) results in a structural interpretation (a rhythmically repeated pair of sounds) whose propositional interpretation is of two objects striking a third. The visual information (patches of colour moving cyclically up and down, with a horizontal component) results in a structural interpretation (a person whose legs are moving). The propositional interpretation is of some person walking in a particular direction. If the sound and picture are synchronised, the output of the acoustic path can be blended with the output of the visual path, to produce a combined representation of 'footsteps'. If there is a cut in this scene to show a close-up of the face of the walking person, then visually any frames could be used: it is not necessary to ensure that the walker is at the same point in a step across the cut, since the face remains fairly steady while they are walking. Acoustically, however, it is crucial to ensure that the rhythm of the footsteps is not disrupted: this will be picked up at the structural level and be interpreted propositionally as the walker missing a step. The viewer will notice this, and the change in the shot will become apparent. In practice, a film-editor can continue the soundtrack from the first sequence over the cut, replacing the soundtrack from the second sequence.

In ICS terms, the viewer is able to continue attending to the scene across the filmic cut at a propositional level, and so the reciprocal loop between the propositional and implication levels of meaning can continue following the action and constructing the narrative. All three of the cyclical flows illustrated in Figure 1 can continue smoothly. If the cut had been unfilmic, then the propositional blending of the information streams would have failed. Immediately following the cut, the output of the propositional subsystem back to the structural level of representation would not have blended with the incoming acoustic information. The viewer would briefly have had to interrupt the higher level comprehension loop to engage in some reciprocal activity between the propositional and structural levels to work out what had occurred acoustically. It is this diversion of processing activity to the structural levels of representation that gives the viewer the sense of a 'perceptual jump' and alerts them to the presence of a bad cut.

5. EXTRAPOLATION TO INTERFACE DESIGN

These and other examples lead us to suggest that many film-editing conventions have the effect of providing thematic continuity at a structural level, allowing propositional and schematic, implicational comprehension of the scene to continue without interruption. With this working hypothesis, we can derive guidance for the domain of interface design, to argue essentially that across screen changes there should be thematic continuity in the user's structural representations to support coherent propositional processing. By this we do not mean simply keeping peripheral elements of the successive screen unchanged (e.g., keeping a window frame or a menu bar in the same place), nor presenting successive screens in a similar 'style' (e.g., a common layout of slots, or dialogue boxes), since these aspects of the display will not be contributing to the user's active propositional representation when the change occurs. The continuity must be with respect to the elements of the display that the user is processing – and the corollary of this is that if it is not clear what element the user will be processing, then they should be directed towards some element that is to provide the 'link' over the screen change.

This suggestion sounds sensible, even obvious, and yet it is rare to find interface designs that have given consideration to the problems users face in reorienting themselves to the display following a screen change. We have heard designers indicate screen objects that are not functionally necessary to a task or a layout as 'a waste of screen real-estate', and 'screen clutter', regardless of the assistance they might be playing in guiding users to the information they are seeking.

We can go further than the original film-editing domain, and argue that in the more interactional domain of HCI, continuity can also apply 'within a shot.' Salient information or objects should not just appear or disappear from the screen, but should make some sort of entrance or exit. Not being limited by the realistic conventions of narrative film, in which objects slide out of frame, interface designers may be able to create many ways of implementing this – even in films, people appear from behind objects and through doors. A rare example of this in a commercial product can be seen in the Macintosh Finder, where windows 'zoom' in and out of their parent folder or application icons.

In one case that we looked at, a tourist information system had been planned to display, using a trackerball and a PC size screen, detailed large-scale views of a locale to help people find places of interest and plan routes. Because the whole area could not be displayed on the screen at the required scale, it was thought necessary to provide a smaller-scale overview. The design problem was how these should be integrated. The first solution had been to split the screen between both scale views, with a highlit area on the small-scale map (on the top two thirds of the screen) corresponding to the area displayed in the large-scale box (the bottom third of the screen). This would require the user to shift their focus of processing between the two areas in order to co-ordinate the motion of the trackball, and in doing so would produce an effect similar to the problem film-makers have of portraying two people conversing in a moving car. When they moved the highlit area in the top part of the screen to the *left*, all of the elements in the lower part of the screen would have scrolled to the *right*.

Various other designs were suggested, involving progressive compression of parts of a large scale map, with a central uncompressed area. All of these designs shared the first design's aim of providing both large scale and small scale views simultaneously. The alternative, of displaying them sequentially, had been rejected because, in part, it required an interaction and a screen change that was thought too confusing for a 'walk up and use' device, even though it would have been a direct parallel of the 'zoom-in' cut used regularly by film-makers. If it had been implemented to work in the same way that many graphics packages operate, with the centre of the small scale view being made the centre of the large scale view, it might well have been confusing. A better solution is to make the element that the user clicks on (or has selected) the centre of the enlarged view, but even this will make the element move from its original screen position to the centre. The filmic solution, given our hypothesis of thematic continuity, would be to ensure that the selected element, which can be assumed to be what the user is actively processing, remains in the same place following the screen change. None of the graphics packages we have access to consistently follow this advice.

This advice also runs counter to the idea that users should be moved between different views by showing them all of the in-between 'frames'. Barfield, Boeve & Pemberton (1993) argue that "It would be better if the system animated the transition between the two viewpoints to give the illusion that they were rapidly moving from the old to the new viewpoint". This 'fill-in animation' takes time, however. Mackinlay, Card & Robertson (1990) describe a technique, based on logarithmic motion, that "helps with orientation by making it fast for the user to zoom out to get orienting views and then to zoom back in". Rather than emphasise speed of motion, film-makers would suggest an immediate cut between the two viewpoints, but one that maintained thematic continuity. In fact, as technology allows interfaces to become more realistic, fast motion back and forth through virtual space may actually be rather disorienting for the user, inducing too effectively the sensation of motion, as do film sequences shot from a roller-coaster.

6. EXTENSION TO MULTIMEDIA

Multimedia demonstrations are admittedly built more to show off how much can be done rather than how things ought to be done, but a typical example will contain several unrelated objects moving around and rotating, while some music is playing, words are appearing and a voice is speaking. These are so overwhelmingly confusing that it is not surprising that many people dismiss the new capabilities as gimmickry, but films are full of moving objects, sound and voiceovers, and some even have subtitles, and they never seem as baffling. The difference is that the simultaneous 'streams of information' are co-ordinated: they are all directed towards some common meaning and can be comprehended by the viewer within a single schematic, implicational representation.

The problems are not restricted to overambitious demonstrations. Where multiple windows are used to show say, pieces of work and video windows of colleagues with whom a user is collaborating, current systems rarely make any attempt at co-ordinating the 'action'. In an example of one such system, we were shown a graph being annotated by someone who had been talking to us from a video window at the lower right of the graphical display. He leant down to draw on 'his' copy of the figure, and his alterations appeared above him and to his left. Structurally this corresponds to a film actor throwing a ball off screen in one direction and it reappearing from another: we could understand it if there *had* been a cut to a different viewpoint (as in Frith and Robson, 1975). Even without a cut, the display would have been structurally coherent if the video window had been above and to the left of the graphical window, so that when our 'colleague' leant down to draw, his marks appeared below him.

7. CONCLUSIONS

In this paper we have shown how a theoretical explanation of craft based skills can allow them to be transferred between domains. We would also argue that this approach allows skills within a domain to be re-used as the domain evolves. While we do not want to discourage people from producing specific guidelines for designers, they should be firmly grounded in a theoretical framework that is independent of the domain. Although the domain may change, the underlying theory remains the same; and while new mappings must be made from the knowledge embodied within the theory to the applied domain, this is less problematic than having to begin research afresh.

ACKNOWLEDGEMENT

This work was carried out as part of ESPRIT BRA 7040 AMODEUS-2, supported by the CEC.

REFERENCES

Adams, W. B. (1977) Handbook of motion picture production. New York; London: Wiley.

Balázs, B. (1970 [1945]) *Theory of Film*. Dover: NY.

Barfield, L., Boeve, E. & Pemberton, S. (1993) Objects, Invariants and Treacle. In J. Alty, D. Diaper & S. Guest (eds) *People & Computers 8*, CUP: Cambridge, 157-171.

Barnard, P. J. (1985). Interacting cognitive subsystems : A psycholinguistic approach to short-term memory. In A. Ellis, (ed.), *Progress in the Psychology of Language, Vol. 2*, 197-258. Erlbaum: London,.

Barnard, P. J., & May, J. (1993). Cognitive Modelling for User Requirements. In P. F. Byerley, P. J. Barnard, & J. May (Eds.), *Computers, Communication and Usability:* Elsevier, Amsterdam.

Barnard, P.J., & May, J. (1995) Interactions with Advanced Graphical Interfaces and the Deployment of Latent Human Knowledge. In F. Paterno' (ed) *Eurographics Workshop on the Design, Specification and Verification of Interactive Systems*, pp. 15-48. Berlin: Springer Verlag.

Bernstein, S. (1988) *The technique of film production.* London: Focal.

Boorstin, J (1991) *The Hollywood eye: what makes movies work*. New York: Harper Collins

Carroll, J.M. (1980) *Toward a Structural Psychology of Cinema*. Mouton: The Hague

Eisenstein, S.M. (1949) *Film Form* trans. J. Leyda. Harcourt Brace: New York.

Frith, U. & Robson, J. E. (1975) 'Perceiving the language of films', *Perception, 4*, 97-103.

Halliday, M. A. K. (1970). Language structure and language function. In J. Lyons, (Ed.), *New Horizons in Linguistics*. Middlesex, England: Penguin.

Koons, W.R., O'Dell, A.M., Frishberg, N.J. & Laff, M.R. (1992) The computer sciences electronic magazine. *CHI'92*. ACM: New York, 11-18.

Long, J. (1989) Cognitive Ergonomics and Human-Computer Interaction. In: J.B. Long and A. Whitefield (eds): *Cognitive Ergonomics and Human Computer Interaction*, CUP: Cambridge, 4-34

Mackinlay, J.D., Card., S.K., & Robertson, G.G. (1990) 'Rapid controlled movement through a virtual 3d workspace', In *Proceedings of Siggraph'90*, 171-176

May, J., Barnard, P.J. & Blandford, A. (1993): Using Structural Descriptions of Interfaces to Automate the Modelling of User Cognition, *User Modelling and User Adaptive Interfaces, 3*, 27-64.

May, J., Tweedie, L. & Barnard, P.J. (1993) Modelling User Performance in Visually Based Interactions. In J.L. Alty, D. Diaper & S. Guest (eds) *People and Computers 8*. CUP: Cambridge. 95-110.

Münsterberg, H. (1970 [1916]) *The Film: A Psychological Study*. Dover: New York.

Pudovkin, V.I. (1958 [1929]) *Film Technique and Film Acting*. Grove: New York

Smets, G., Overbeeke, K. & Gaver, W. (1994) Form Giving. *CHI'94*. ACM, NY. 79-84.

Teasdale, J. & Barnard, P.J. (1993) *Affect, Cognition & Change*. LEA: Hove

Tognazzini, B. (1993) Principles, Techniques and Ethics of Stage Magic and their application to human interface design. *InterCHI'93* . ACM, New York, 355-362.

Young, E. & Clanton, C. (1993) 'Film Craft in User Interface Design'. Tutorial at *InterCHI'93*, Amsterdam.